# FINAL THESIS REPORT

## Designing preferable diets using mathematical modeling: A review of preference elicitation methods


**Yohana Putri Nareswari**
**900913596050**

**Food Quality Design (FQD 80436)**

**Supervisors :**
Argyris Kanellopoulos (Operations Research and Logistics group)
Ante Ivancic (Operation Research and Logistic group)
Elsbeth Spelt (Food Quality Design)

# ACKNOWLEDGEMENT

The completion of this thesis could not have been possible without the assistance and supports of so many people whose names might not be all enumerated. Firstly, I would like to express deep appreciation to the following:

My first ORL supervisor, Argyris Kanellopoulus, for his patience, motivation, and support in guiding me through the research process. Despite his busy schedule, he always tried to allocate a discussion time for me (plus a free cup of hot chocolate) almost every week. His kindness is indeed something that I am grateful for.

My second ORL supervisor, Ante Ivancic for his knowledge, input, and encouragement to finish my research. His expertise in machine learning greatly helped me to understand machine learning techniques.

My second FQD supervisor, Elsbeth Spelt for her suggestions and critical point of view that helped me a lot to improve my report and my research writing skill. Without the assistance of ally my supervisors, I would not be able to conduct and finish this research on time. They are indeed many aspects that I still have to improve, but I learned so much from this learning process.

My parents and my partner for their unconditional love and unlimited support that motivated during the 'ups' and 'downs.'

My FQM and thesis colleagues for supports and friendship. I am thankful that we could share our difficulties, motivate each other's, and struggle together! Hopefully, this friendship can last forever!

The last but not the least, to the Indonesian Endowment Fund for Education (LPDP) for financing my study at Wageningen University.

Wageningen, March 2017

Yohana Putri Nareswari

# TABLE OF CONTENTS

# ABSTRACT

Nowadays, the world faces greater challenges in sustaining resources than before. People might be aware that food is essential for their health, but the impacts of food production and consumption on world's resources are less well known. The food system challenges trigger the need to modify the existing consumption pattern to be healthier, more sustainable, as well as affordable. Aligned with those issues, European Commission is collaborating with a number of institutions to conduct a research program called SUSFANS which aims to improve EU agri-food sectors. The modelling of SHARP diets for European Union (EU) consumers is one of the targets. SHARP stands for the diet criteria, namely Sustainable, Healthy, Affordable, Reliable, and Preferable.

Diet modelling is a decision support tool that helps to determine the optimal diet based on several predefined indicators. The alteration of dietary behaviour to either improve health or reduce GHGEs is often perceived as unfavourable because they do not to take into account consumers 'actual preferences and the reasons behind their food choices. Thus, it is important to include the preference of decision maker (DM) to generate a sensible optimized diet. This would imply that somehow we need to recover or elicit preferences of the DM. There is a number of existing methods to elicit DMs' preferences. Therefore, this research aimed to review the existing typical preference elicitation methods and identify under which conditions those methods could be applied in SHARP diet model.

Prior to the actual literature review, a preliminary literature review was conducted to build a conceptual review framework that provides a step by step guideline to review the preference elicitation methods. There were two literature search strategies used for this research, namely snowball method and systematic search. The main search strategy was the snowball method which collected relevant publications from several leading publications. Initially, the preference elicitation methods were collected from those publications and classified into two main categories (compositional and decompositional) under several sub-categories:

(i) **Compositional**: *direct rating (i.e. scales, point allocation, SMART, interval SMART); ranking (i.e. direct ranking, SMARTER); pairwise comparison (i.e. AHP, MACBETH); swing weighting (i.e. SMARTS); and scoring function (i.e. bisection))*

(ii) **Decompositional**: *choice-based (PAPRIKA, BWS, CBCA/DCE); rating-based (RBCA); and learning-based (i.e. collaborative filtering, PageRank).*

Those methods originate from the major applications, namely multi-criteria decision analysis (MCDA) (i.e. compositional methods), conjoint analysis (i.e. PAPRIKA, BWS, CBCA/DCE, RBCA), and machine learning (i.e. collaborative filtering, PageRank). Subsequently, the procedures, main components, as well as the advantages and disadvantages of preference elicitation methods were identified. The description of each method was accompanied by illustrative examples in diet model-context to provide a better overview. Based on the systematic search, the trends from 2001 to 2016 indicated that decompositional methods were more frequently used than compositional methods. Those reviewed methods were applicable in SHARP diet model, depending on several conditions. The conditions are *the type of data collection (interactive or empirical); the SHARP diet criteria (limited or all); the aggregation level of the consumer (individual(s) or average-individuals); and dealing uncertain data (deterministic or stochastic)*. In general, MCDA methods could be applied in almost similar conditions as conjoint analysis methods, except in terms of dealing with data uncertainty. The conditions for machine learning methods mostly differ from both methods. In conclusion, there are various methods to elicit preference of DM which vary in several aspects. Furthermore, they could be applicable in the development of SHARP diet model.

## LIST OF ABBREVIATIONS

| | |
|---|---|
| AHP | Analytical Hierarchy Process |
| ACA | Adaptive Conjoint Analysis |
| BWS | Best worse scaling |
| CA | Conjoint Analysis |
| CBCA | Choice-based conjoint analysis |
| CF | Collaborative Filtering |
| CI | Consistency Index (in AHP) |
| CM | Conjoint Measurement |
| $CO_2$ | Carbon dioxide |
| CR | Consistency Ratio |
| DCE | Discrete Choice Experiments |
| DM | Decision Maker |
| EU | European Union |
| FNS | Food and Nutrition Security |
| GHGEs | Greenhouse Gas Emission |
| MACBETH | Measuring Attractiveness by a Categorical-Based Evaluation Technique |
| MAUT | Multi-attribute Utility Theory |
| MAV | multi-attribute valuation |
| MAVT | Multi-attribute Value Theory |
| MCDA | Multi-Criteria Decision Analysis |
| MFE | Meat/Fish/Eggs |
| ML | Mixed Logit |
| MP | Mathematical Programming |
| OLS | Ordinary Least Square |
| PAPRIKA | Potentially All Pairwise Rankings |
| RBCA | Rating-based conjoint analysis |
| RE | Rank Exponent |
| ROC | Rank of Centroid |
| RP | Revealed Preference |
| RR | Rank Reciprocal |
| RS | Rank Sum |
| RSs | Recommender Systems |
| RUT | Random Utility Theory |
| SHARP | Sustainable, Healthy, Affordable, Reliable, and Preferable |
| SMART | Simple multi-attribute rating technique |
| SMARTER | SMART exploiting ranks |
| SP | Stated Preference |
| SQ | Sub-questions |
| TTO | Time-Trade Off |
| WLS | Weighted least squares |

# CHAPTER I. INTRODUCTION

## *1.1 Background of the research*

### *1.1.1 Food system's challenges*

Nowadays, the world faces greater challenges in sustaining resources than before. People might be aware that food is essential for their health, but the impacts of food production and consumption on world's resources are less well known. It is estimated that the world's population will increase to 9.2 billion people in 2050 or around 34 percent higher than today (Freibauer *et al.*, 2011). Nevertheless, the available resources (such as water and land) are increasingly limited. Therefore, to sustain the resources, the concern of diet is shifting to go further beyond 'good for health' to 'good for the planet' (European Commission, 2015).

Globally, including in Europe, many of the food production systems are surpassing the environmental limit or are almost doing so. For example, nitrogen synthesis exceeds the planetary limit by a factor of four, while the phosphorus use has touched its global boundary (European Commission, 2015). It was approximated that the current food system contributed to about 25% of total greenhouse gas emission (GHGEs) (Sjörs *et al.*, 2016). Thus, reduction in GHGEs emission by 2050 of 50% globally is required to avoid climate change (Millward and Garnett, 2010).

The food systems emit GHGEs at all stages of its life cycle, from farming to manufacturing, distribution, storage, consumption, and waste disposal. From those steps, it is gradually observed that livestock's contribution to the total food burden is significant, about 18% of GHGEs (Kingston-Smith *et al.*, 2010). The whole supply chains of animal-based diets contribute to the highest GHGEs production (Millward and Garnett, 2010), it was estimated that the contribution accounts for 13% of all European Union's GHGEs (half the total impact of food) (Guinée *et al.*, 2006). However, there are also large variations of GHGEs within the groups of meat and dairy products (Sjörs *et al.*, 2016). For example, the carbon dioxide $CO_2$ emission for one kg of beef is almost 48 kg while for one kg of poultry is about 4 kg $CO_2$ (Sjörs *et al.*, 2016).

The food system challenges above trigger the need to modify the existing consumption pattern to be healthier, more sustainable, as well as affordable. In order to meet the reduction target of GHGEs, the changes in animal-based diet are promoted. Even though diets rich in animal sources (i.e. meat and dairy) are nutritionally high in protein, calcium, iron, and vitamin B12; the diet that is lower in meat intake is believed to promote healthier life by decreasing the risk of obesity, hypertension, and cancer as well as lower environmental footprint (European Commission, 2015).

### *1.1.2 SHARP diets*

Diet itself is defined as '*food and drink regularly provided or consumed by a person*' (Merriam-Webster, 2016). Diet modelling is a decision support tool that helps to determine the optimal diet based on several predefined indicators (Day *et al.*, 2008). Aligned with those issues, European Commission is collaborating with a number of institutions to conduct a research program called SUSFANS which aims to improve EU agri-food sectors. The modelling of SHARP diets for European Union (EU) consumers is one of the targets. SHARP stands for five dimensions, namely Sustainable, Healthy, Affordable, Reliable, and Preferable (SUSFANS, 2015). **Sustainability** implies 'the use of resources at rates that do not exceed the capacity of the Earth to replace them' (European Commission, 2015). **Healthy** means nutritionally sufficient while **affordable** refers to food that can be afforded by all socioeconomic group, yet still supporting the EU agri-food sector. **Reliable** indicates stability in food supply while **preferable** suggests consistency with cultural norms and preferences (SUSFANS, 2015). The SHARP diets are expected to deliver options for sustainable food and nutrition security (FNS) which can be applied across EU. Hence, the food system challenges above could be solved.

According to the previous research on the review of current diet modelling, each dimension of SHARP diet has performance indicator(s), except reliability dimension (Faramitha, 2016) as presented in Table 1. The author reasoned that it was because logistical matters might be seemed to be not directly related to consumer's diet. GHGEs is considered as the most representative indicator of sustainability because GHGEs were calculated in whole food chains and mainly represented as $CO_2$ emissions. The health dimension was evaluated from the nutrient and the energy contents of the diets. The indicator for affordability was diet cost, while preferability dimension was indicated by the total deviations from current diets, food intake, food inclusion or food exclusion, and number of servings. The food intake (usually expressed in gram/day) can be the indicators for health and preferability dimensions, but the result indicated that the food intake was commonly used as preferability constraint. The limit can be established from the observed food intake or dietary guidelines. Consequently, it was expected that the diets might have a high acceptance since the suggested food choices were the food that people usually consume (Faramitha, 2016).

*Table 1. SHARP dimensions and indicators (taken from Faramitha (2016))*

| SHARP DIMENSION | THE SELECTED PERFORMANCE INDICATOR(S) |
|---|---|
| **SUSTAINABILITY** | GHGEs or mainly $CO2$ emission |
| **HEALTH** | Nutrition content, Energy content |
| **AFFORDABILITY** | Diet cost |
| **RELIABILITY** | - |
| **PREFERABILITY** | Deviation from current diets, food intake, food group, number of servings |

### *1.1.3 Decision makers' preferences in diet modelling*

Most of the diet modellings found in the literature were focused on nutrition adequacy (Maillot *et al.*, 2010) and combination between nutrition adequacy as well as sustainability (Horgan *et al.*, 2016). Horgan *et al.* (2016) stated that the alteration of dietary behaviour to either improve health or reduce GHGEs is often perceived as unfavourable by the general public, even though there is support for the healthy and sustainable diets. The major limitation of many dietary guidelines is that they do not to take into individuals food choices and habits which driven by preferences (Horgan *et al.*, 2016). Therefore, these theoretical diets cannot be adopted by all individual, since the diet with minimum GHGEs might eliminate some favourable food groups, specifically Dairy and Meat/Fish/Eggs (MFE) (Perignon *et al.*, 2016) and makes nutrients harder to get (Raffensperger, 2008).

In a SHARP diet, this diet modelling might recommend what type of foods or drinks needs to be consumed to meet nutrition, affordability, reliability, and preferability requirements by addressing the sustainability issues. There are some decisions to be made. Therefore, it is important to include the preference of decision maker (DM) to generate a sensible optimized diet. This would imply that somehow we need to recover or elicit preferences of the DM. In an example from Ribal *et al.* (2016), the preference of DM was depicted by the preference weights. These preference weights denote the relative importance of each attribute/ criteria for DM. The interactions of those factors and their relative importance for DM might improve the applicability of the resulted diet recommendation. There is a number of existing methods to elicit DMs' preferences which also can be useful for the SHARP diet model. Therefore, it is important to review those preference elicitation methods.

<u>Demarcation of the study</u>

There are multiple methods which can be applied to incorporate preferences of individuals to the models. These techniques are called as *preference elicitation methods*. The techniques can be completed by following different approaches: (a) by using query interface where individuals are requested to express their preferences, or (b) by capturing implicit individuals' choices and motives, as well as employing data/preference mining algorithms (De Amo *et al.*, 2015).

Different preferences elicitation methods have been known to generate different results. The chosen elicitation method matters because they diverge in the amounts and type of acquired information, data collection effort, and method of analysis (Carson and Louviere, 2011). Moreover, those preferences elicitation methods are barely used in diet modelling since most models mostly focus on minimizing the departure from the observed diet or using equal preference weights in the diet model. Due to those factors, it is intriguing to gain knowledge upon the features of different preference elicitation methods as shown in Figure 1. Figure 1 below depicts the proces of methods selection in which the review of various methods' attributes plays a great role. Thus, this study is valuable to address the gap of knowledge and to enhance the growing research of diet modelling.

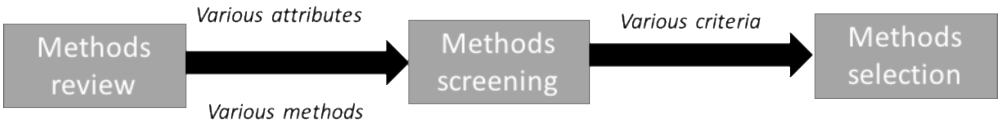


*Figure 1. The process of method selection*

## 1.2 Problem statement

Using preference elicitation methods will assist in improving the existing theoretical diet models by accurately modelling the individuals' preferences. The inclusion of individual preferences into the diet modelling is expected to improve predictive capacity and trust of policy makers on model's outcome. This will enable us to assess future scenario based on the forecast of how an individual will adapt their diet if certain factor changes in the future.

There are multiple methods available for such objective. Those methods have been applied widely in various disciplines, but not in diet modelling. They might differ in several aspects, such as complexity, applicability, data requirements, and results. Moreover, the advantages and disadvantages of those approaches are not fully well-known. Thus, it is useful to review the existing methods in order to comprehend those different approaches and then be able to use them in diet modelling.

## 1.3 Research objectives

The objective was to review the existing typical preference elicitation methods and identify under which conditions they can be applied in diet modelling. This study will provide more insight into how to determine the appropriate preference elicitation methods for SHARP diet modelling. Further, this could be supporting knowledge for developing SHARP diet modelling.

## 1.4 Research questions

**Main research question**: What are the typical preference elicitation methods used in current decision support tools as being reported by scientific literature and under which conditions they can be appropriately applied in SHARP diet modelling?

**Sub-questions**:
1. What are the steps of those methods?
2. What are main differences and features of existing preference elicitation methods?
3. What are the advantages and disadvantages of each method?
4. What are current trends in preference elicitation methods in decision making-related application?
5. When can the methods be appropriately applied in diet modelling and under which conditions?

### *1.5 Research approach*

The research approach is depicted in Figure 2. Firstly, a preliminary literature review was conducted to comprehend the application, the characteristics, and the classification of preference elicitation methods. The following questions were used for the preliminary literature review:

1. How do people make a choice?
2. What is preference elicitation method?
3. How are preference elicitation processes conducted?
4. What are the characteristics of preference elicitation methods?
5. In which fields the preference elicitation methods are usually applied?
6. How the preference elicitation methods are usually classified?
7. What are the main characteristics for each main category?
8. What are the important parameters for comparing preference elicitation methods?

This information was helpful to build a conceptual review framework because it provided insights on which important information that should be obtained related to preference elicitation methods. The conceptual review framework was then used as a guideline to conduct the literature review.
The main method used for literature review was the snowball technique. Initially, a number of interesting publications were chosen as leading publications. In order to identify the initial set of leading articles, it was important to specify some criteria to ensure the reliability, relevance, and accuracy of their contents. These main articles were chosen based on some criteria: (i) they should be preferably review papers that provide the state of the art in certain research field; (ii) they should be recently published; (iii) highly cited; and (iv) in highly ranked journals in the field of operational research and management science/consumer studies/ health. Then, a set of relevant publications from the reference lists and the publications that have cited these leading publications were tracked down. Those relevant publications were then used to answer sub-questions (SQ) 1, 2, 3, and 5.

Subsequently, we also identified several keywords from those publications. These keywords were used for systematic search using Scopus and Web of Science databases. The systematic search assisted in answering SQ 4, which was related to the trends in preference elicitation methods.

The main research question was answered after solving the SQs. The gathered publications provided the list of the common preference elicitation methods. The conceptual review framework provided the guideline of activities that should be conducted. The methods were then classified according to the classification from the preliminary literature review. There were several different classifications of the elicitation methods from the review papers. The most representative classification was chosen and applied to this research. Additionally, the publications gave an extensive description of the preference elicitation methods which included their procedures and features. This present research did not elaborate all the methods exhaustively, but it focused on providing a basic understanding of the preference elicitation methods with illustrative examples in diet modelling context. The methods were compared among each other by summarizing the important parameters and main components. From that information, the advantages and disadvantages of preference elicitation methods can be derived. The main research question was answered after analysing under which conditions the reviewed preference elicitation methods are appropriately used in the SHARP diet model.

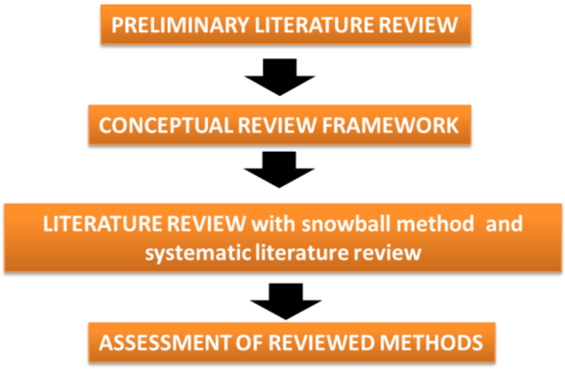


*Figure 2. The research approach of this present research*

The subjects of the successive chapters are briefly described as following:

Chapter 1 proposes and explains the overall research idea by emphasizing the importance of incorporating individuals' preference on diet modelling and choosing appropriate preferences elicitation methods for SHARP diet modelling. The research objective, research questions, and research approach helped are also explained to provide clear direction.

Chapter 2 provides the general information about preference elicitation methods and diet modelling from the preliminary literature review. The obtained information was used to build a conceptual review framework. This conceptual review framework works as the guideline and contains the activities that should be done during literature review.

Chapter 3 covers the research methodology used for this research. This chapter explains the steps of the literature review for snowball method and systematic literature search. This chapter includes the identification process of relevant publications, the extraction process of relevant data, and the description of data analysis activity.

Chapter 4 presents and discusses the results of the research. The results comprised of the classification of the obtained methods, the procedure, and description with examples in diet model-context. This chapter also covers the evaluation of advantages and disadvantages of the reviewed methods and under which conditions they could be applied in SHARP diet model.

Chapter 5 concludes the overall research results regarded to the research objectives and questions stated in Chapter 1. This chapter also discusses several limitations of this research and recommendations for the future research.

Chapter 6 contains personal reflections together with the insights and the obstacles of the research process.

# CHAPTER II. PRELIMINARY LITRATURE REVIEW

This chapter presents the information from the preliminary literature review and consists of three main sections. The first section focuses on background information about preference elicitation methods. This consists of definition, applications, classification, and characteristics of preference elicitation methods. Information regarding diet modelling is explained in the second section. The third section summarizes the key points of the obtained information in a conceptual review framework. A conceptual review framework aims to explain the major issues to be studied, including key factors, concepts, or variables and their presumed relationship (Miles and Huberman, 1994). This framework provides a guideline to conduct the review of preference elicitation methods.

## *2.1 Preference Elicitation Methods*

The preference elicitation method is related to how people make choice. In order to provide a clear explanation, the choice process is discussed briefly. It is followed by elaborated definition of preference elicitation methods and their applications. Then, the important parameters and main components of preference elicitation methods are also included due to their significance in comparing the methods.

### *2.1.1 Choice behaviour*

People tend to choose certain diet pattern among various available diet alternatives. Thus, diet preference is an example of choice behaviour. Human choice behaviour is regarded as "*a mental process that transforms perceptions of several optional courses of action into a choice. It is considered to cover any kind of intuitive, automatic and impulsive choice behaviour as well as conscious-deliberate decision making"* (Van De Kaa, 2010). The choice behaviour process can be considered as a system that transforms inputs into outputs within an environment (Van De Kaa, 2010). **The inputs** of each individual choice process are insights of the individual's choice context and of her/his concurrent needs**; the outputs** are then the choices. *The choice context* comprises of the environment (the '*state of the world'*) and the '*state of the organism*' (concurrent moods, needs, belief, etc.).

Figure 3 depicts the framework of human choice behaviour. The goal of choice behaviour process is to select one feasible course of action from a set of alternatives that, in that specific context, satisfies individual's concurrent needs. There are four mental functions related to choice behaviour process in subject's minds (Van De Kaa, 2010):

i. **The framing of the choices and needs by mental perception** into some *choice alternatives* (i.e. representations of possible courses of action and their expected outcomes in terms of probabilities and attributes) and *preferences* correlated to the choice subject's concurrent needs, desires, as well as goals

ii. **Judgment** involves *the assessment of the sizes/characteristics* of the expected outcomes (attributes and probabilities) and *valuation* of these probabilities and attributes' characteristics.

iii. **Evaluation-and-choice** includes an *evaluation* of relevant outcomes from the alternatives and *the selection* of the alternative that meets the criteria.

iv. **Choice behaviour strategy** is required to coordinate the process.

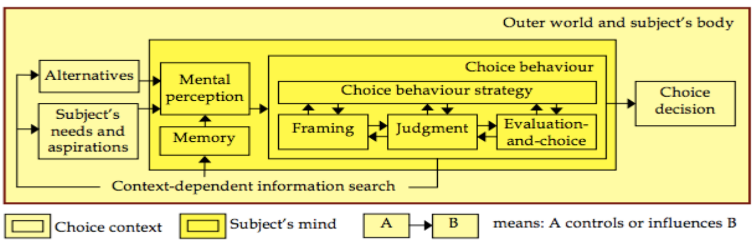


*Figure 3.Framework of choice behaviour (Van De Kaa, 2010)*

### *2.1.2 Definition of Preference Elicitation Methods*

In making choices, economists assume that individuals choose the alternative that delivers them with the greatest utility (Ali and Ronaldson, 2012). The utility itself can be defined as "satisfaction" (Ali and Ronaldson, 2012), "*the quality or state of being useful"* (Merriam Webster, 2016), or "*the value a decision-maker relates to a certain outcome*"(Riabacke *et al.*, 2012). In the latter definition, preferences are associated with a certain number or *value* $(V(i))$ for each alternative $i$ by taking all criteria into account (Belton and Stewart, 2002). The utility of each choice alternative is not observed by researchers, however the choice decisions made by individuals can reveal the underlying utility or value they associate with each alternative (Ali and Ronaldson, 2012). Preference elicitation methods uncover systematic components that govern people's evaluations of objects (Huber *et al.*, 1993). Thus, the preference elicitation methods are useful in revealing the utilities or values of alternatives and the corresponding criteria.

The methods might incorporate graphical, numerical, and verbal expression to help the decision makers in revealing their subjective values of competing criteria or alternatives (Riabacke *et al.*, 2012). Riabacke *et al.* (2012) explained three types of information that can be derived from user input, namely probabilities, utilities, and weights. Probability information is normally elicited from domain experts or from learned data, whereas utility is employed to reflect decision-makers' individual risk behaviour accurately, and weights reveal the importance of one dimension relative to others in terms of scores (Riabacke *et al.*, 2012). However, they are highly correlated to each other. Therefore, those three types of information will be used interchangeably in this study.

### *2.1.3 Application of Preference Elicitation Methods*

Preference is an interdisciplinary topic that can be learned from different perspectives (Domshlak *et al.*, 2011). Hence, preference elicitation methods can be broadly applied to various groups of the data analysis. After conducting the preliminary literature review, we decided to focus our study on the application of preference elicitation methods related to MCDA, machine learning, and conjoint analysis. The selection was due to (i) the preference elicitation methods widely used in those groups (ii) the consideration that those three groups originate from different analytical approaches. MCDA mainly uses simple approaches (i.e. multi-attribute value), conjoint analysis utilizes statistical analysis, and machine learning is using automatic learning algorithm.

In multi-criteria decision analysis (MCDA), the relative importance of criteria is usually indicated by **weights**, while **alternatives** are evaluated on the weights of the corresponding criteria (Wang *et al.*, 2009). According to Wang *et al.* (2009), MCDA is a decision support approach that is suitable for examining complex problems with conflicting objectives and multi-interest problems. These methods have been broadly utilized in social, economic, agricultural, industrial, ecological, sustainability issues and biological systems (Wang *et al.*, 2009, Cinelli *et al.*, 2014, Ananda and Herath, 2009). Many other terms are used to describe MCDA, such as MODM (multi-objective decision making), MADM (multi-alternative decision making), and MCDM (multi-criteria decision making) (Khalili and Duecker, 2013). According to Riabacke *et al.* (2012), Multi-attribute Value Theory (MAVT) is one the most widely used MCDA methods in practice. Figure 4 describes four main stages involved in MCDA: *alternatives' formulation and criteria selection*, *criteria weighting*, evaluation, and *final treatment and aggregation (Wang et al., 2009).*

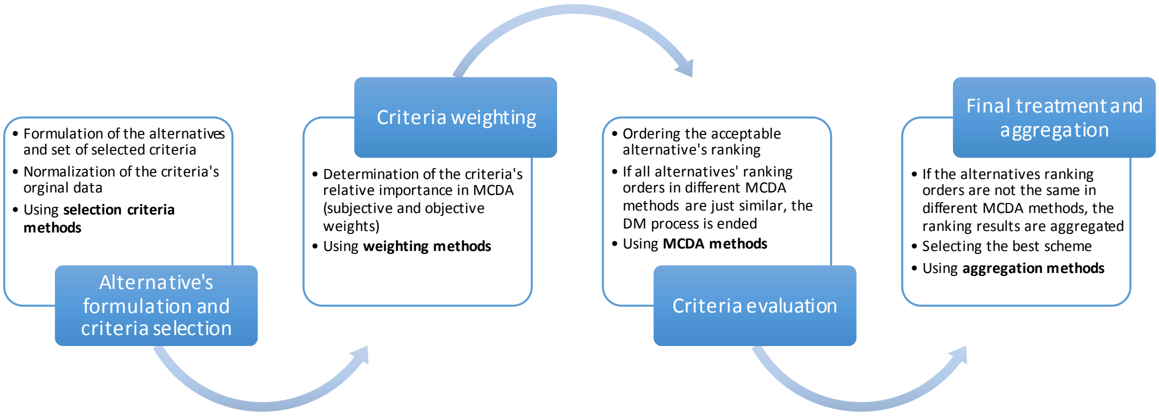


*Figure 4.Four stages of multi-criteria decision analysis (MCDA)*

Fürnkranz and Hüllermeier (2011) observed that preference learning has been growing in artificial intelligence, such as in machine learning. In machine learning, the utility functions are automatically learned from given training data (i.e. recommender system) (Domingos, 2012). Preference elicitation is also widely applied in market research. Conjoint analysis is a well-known family of techniques which are frequently used in market research in order to assess consumer's preference.

### *2.1.4 Classification of Preference Elicitation Methods*

The focus of this section is to present existing categorizations of preference elicitation method. The sources were mainly review papers from numerous subjects that extensively applied preference elicitation methods, namely *health care economics* (Thokala *et al.*, 2016, Marsh *et al.*, 2016, Ali and Ronaldson, 2012), *multi–criteria decision making* (Riabacke *et al.*, 2012, Yusop, 2015),*artificial intelligence and machine learning* ((Domshlak *et al.*, 2011, Kotsiantis *et al.*, 2007), *environmental economics* (Whitehead *et al.*, 2008, Alriksson and Öberg, 2008). According to these papers, there are two possible categorizations: based on the source of information and cognitive process.

a. <u>Based on the source of information</u>

Preference elicitation methods can be divided into two main categories, namely **stated preference elicitation methods (SP)** and **revealed preference methods (RP)** (Whitehead *et al.*, 2008, Thokala *et al.*, 2016, Ali and Ronaldson, 2012, Alriksson and Öberg, 2008, Marsh *et al.*, 2016). Both differ in the sources of information.

**SP** elicitation methods *ask individuals to reveal their preference for a set of two or more alternatives in a* ***hypothetical*** *scenario with different levels of criteria* (Ali and Ronaldson, 2012, Whitehead *et al.*, 2008) or *for a set of criteria* (Thokala *et al.*, 2016, Marsh *et al.*, 2016). For example, the DM is asked to select a diet alternative from a set of prototype diets which have a varying level of numerous criteria. On the contrary, **RP methods** require the *evaluation of* ***actual consumer behaviour*** *(observed) in a real-life market setting (Whitehead et al., 2008, Ali and Ronaldson, 2012)*. For instance, an analysis of actual transactional data to evaluate individual choices of diets in real life after the certain policy has been applied over certain time. Two publications (Alriksson and Öberg, 2008, Ali and Ronaldson, 2012) discovered that revealed preferences are not applicable in a field where actual situations, product, or policy do not exist yet. Meanwhile, stated preference methods are highly recommended when the product, service, or policy is not currently existing in real life, or if the aim is to evaluate preferences among currently available alternatives product (Ali and Ronaldson, 2012).

b. Based on the cognitive process

Based on the process of thinking (so-called *cognitive process)*, the preference elicitation methods can be categorized into ***compositional*** and ***decompositional*** methods (Hanisch, 2012, Green and Srinivasan, 1978). However, some publications also categorised compositional and decompositional approaches under stated preference method (Marsh *et al.*, 2016, Thokala *et al.*, 2016, Alriksson and Öberg, 2008). The first two publications investigated complex health care decision involving multiple-criteria decision analysis (MCDA), whereas the latter focused on the application of conjoint analysis for environmental evaluation. In general, both approaches have opposite pathways. ***Compositional*** approaches evaluate each criterion's utility separately and build up the overall utility of an alternative (Alriksson and Öberg, 2008). As an illustration, the DM can be asked to give scores to each criterion within the diet alternatives (i.e. Direct rating). The best alternative will be the one with *the highest sum of utility*. On the other hand, ***decompositional*** approaches examine the overall utility of alternatives as a whole, from which utility for criteria are derived (Alriksson and Öberg, 2008). In the case of weight elicitation, the utility refers to scores and weights (Marsh *et al.*, 2016, Thokala *et al.*, 2016). For instance, the decision maker is asked to choose the best diet alternative out of the set of diet options, given the performance level of the criteria. This example is referred to Discrete Choice Experiments (DCE). Figure 5 below summarizes the definition of compositional (orange boxes) and decompositional (green boxes).

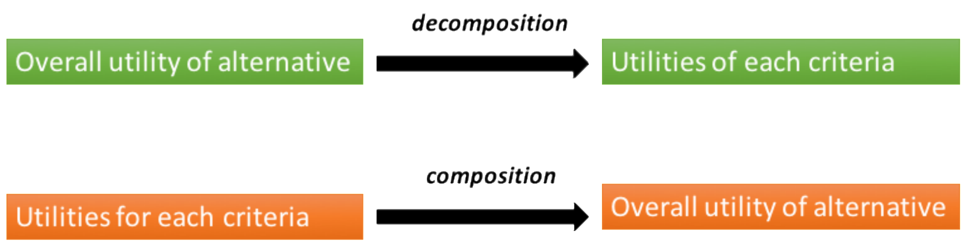


*Figure 5.The mechanisms of compositional and decompositional methods*

In the context of diet modelling, the motives that play roles affect food choices can be considered as criteria. Some motives for food choices can be sensory, availability, price, health, mood, familiarity, convenience, ecological, political, region, and naturalness (Honkanen and Frewer, 2009). The definition of **revealed preference** might be overlapping with the ***decompositional*** approach since observed data also contain the overall value of alternatives which will then be decomposed into weight or scores of criteria.

Some publications even categorized ***decompositional*** and ***compositional*** under SP approaches, but no further information regarding RP approaches found in the literature (Alriksson and Öberg, 2008, Marsh *et al.*, 2016, Thokala *et al.*, 2016). Thus, in order to avoid confusion and misconception, we prefer to use ***compositional*** and ***decompositional*** terminologies for the rest of this report because they are easier to distinguish and covering the aspect of SP and RP approaches.

There are various types of preference elicitation methods (i.e. absolute measurement, pairwise comparisons, ordinal judgments, rankings, choice) (Aloysius *et al.*, 2006). For this research, the **decompositional** and **compositional** classification will be adapted from the publication of Marsh *et al.* (2016) as illustrated in Figure 6. The article was selected because it provides a structured classification and references to some related journals. In the original classification, the ranking-based and learning-based sub-categories did not exist. These categories were added to accommodate rating-based conjoint analysis method and machine learning methods respectively. The following part displays the sub-categories of compositional and decompositional categories. The comprehensive description of the methods within every sub-category is presented in Chapter 3.

***Compositional method**: eliciting utility (preference) of each criterion, and then deriving the utility of alternatives by composing the utility of all criteria.*
***Decompositional method**: eliciting utility (preference) for each alternative, then deriving the utility of each criterion by decomposing the utility of alternative (Marsh et al., 2016, Alriksson and Öberg, 2008, Huber et al., 1993, Green and Srinivasan, 1978)*

**The compositional methods are divided into five sub-categories:**

a. *Ranking* is considered as the simplest approach for assigning weights. Principally, the criteria are ranked in order from most important to least important (Yusop, 2015).
b. *Direct rating* requires the DM to assign points indicating the criteria weights directly. The DM is asked, for example, to distribute 100 points among the criteria. The higher point a criterion has, the greater its relative importance (i.e. point allocation) (Yusop, 2015).
c. *Pairwise comparison* is made by comparing alternatives pairwise on each criterion (Marsh *et al.*, 2016). Their "intensity of importance" relative to each other is usually expressed on "semantic scale" (Yusop, 2015). The example is Analytical Hierarchy Process (AHP).
d. *Swing weighting* uses the "swing" to denote the relative importance of ranges of performance on each criterion (Marsh *et al.*, 2016).
e. *Scoring function* defines the score that will be credited to all levels of performance along with a criterion. It can be generated using bisection approaches. The advantage is that the relationship between the performance on a criterion and preference for that performance is transparent (Marsh et al., 2016).

**The decompositional methods are divided into:**

a. *Choice-based* is applied respondents make choices from or rank a series of sets of product/ diet profiles (Asioli *et al.*, 2016). Some examples are best-worse scaling (BWS) and choice-based conjoint analysis.
b. *Rating-based* used when respondents rate their preference for different product profiles (alternatives) (Asioli *et al.*, 2016). An example is rating-based conjoint analysis.
*c.* *Learning-based* refers to machine learning technique which utilizes learning-algorithm (Kotsiantis *et al.*, 2007). This sub-category is an additional to the model of Marsh *et al.* (2016). This technique uses a set of training data containing items (e.g., diet alternatives) for which preferences are known. The task is to learn a function that predicts preferences for a new set of items (i.e. new diet alternatives), or for the same set of items in a different context (i.e. the same diet but for a different user) (Domshlak *et al.*, 2011).

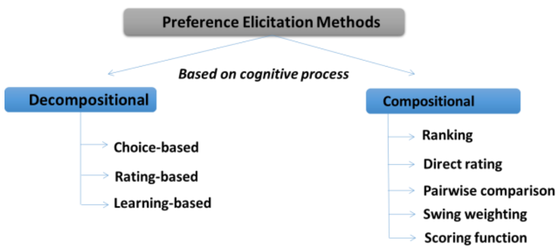


*Figure 6.The classification of preference elicitation methods (adapted from (Marsh et al., 2016)*

### *2.1.5 Preference Elicitation Process*

In the following section, we will present a simple flow of preference elicitation process adapted from Nikou *et al.* (2015). Within the process, there are two major actors: (1) *the researcher* or *analyst* (choose and translate the method to a feasible process for respondent, as well as analyse the obtained information); (2) *the consumer or respondent* (express preferences in the choice process). Figure 7 describes preference elicitation process which consists of *method selection, data collection*, and *data analysis* steps in order to understand the consumers' preferences.

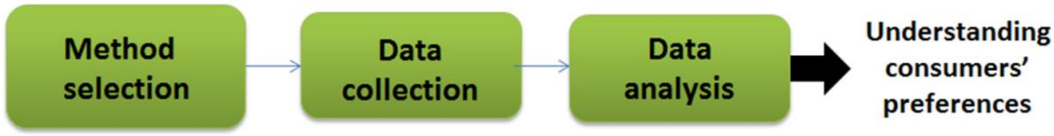


*Figure 7.Preference elicitation process and goal*

Initially, the analyst decides which preference elicitation method to be used based on the understanding of the problem, information about respondents, and his/her knowledge of the methods. Subsequently, the analyst designs the data collection process to be performed. The data collection can be conducted by gathering respondents' judgements through surveys (i.e. conjoint analysis) or collecting transaction database (i.e. machine learning). Afterwards, the collected information will be analysed. Based on the analysis, the analyst translates the result and derives a conclusion regarding consumers' preferences.
Therefore, the chosen preference elicitation method must be well-understood by the analyst. The respondent also needs to comprehend how the questionnaire works and what the problem is. In this case, the respondent should know that they have to give their preference over some food attributes or diet alternatives. If the selected method is too complex and not comprehensible, then the options are either to provide additional training or change to an easier method.

### 2.1.6 Characteristics of Elicitation Methods

Discovering the suitable methodology to elicit individuals' preference is an essential step for the subsequent success or failure of a diet. Since this study aims to collect the methods and examine their capabilities of supporting a diet model, then the characteristics (*inherent properties*) of the methods are important to understand their applicability (Marsh et al., 2016). These characteristics consist of two major parts, namely *main components* and *important parameters*. These characteristics are advantageous to be used to compare the methods.

a. **Main components**

Riabacke *et al.* (2012) stated that the elicitation methods have distinct features which influence their applicability in practice. Therefore, those features need to be addressed more explicitly. For that reason, Riabacke *et al.* (2012) categorized the main components of the elicitation into three main components as shown below:

(1) **Extraction** considers how information (weights, probabilities, utilities) is derived through the DM's input. It deals with the type of survey questions (i.e. asking rate from conjoint stimuli, ratio-scale from pairwise comparison, etc.) or source of input (i.e. purchase history). *Ratio-scale* means that the respondent assigns a certain value on that criterion in comparison to another criterion.

(2) **Representation** relates to the format of representing the DM's input so that it could capture the retrieved information (i.e. point estimate, semantic estimates, ordinal, etc.). Point estimate refers to cardinal values, such as ratio-scale; while semantic indicates categorical group, such as "2 means good, 1 means intermediate".

(3) **Interpretation** deals with how to assign meaning to the captured information on the evaluation of the alternatives. The interpretation is divided into two aspects.

- **Utility model:** The model corresponds to how the final utility is calculated (i.e. weighted additive utility model, additive part-worth utility model) (Nikou *et al.*, 2015).The additive part-worth utility usually applies to conjoint analysis methods.

*Part-worth* mostly used in the conjoint analysis. This term refers to the estimated utility of each criterion after decomposition of utility alternative (Alriksson and Öberg, 2008).

- **Procedure** : The procedure relates to how they evaluate the obtained data (i.e. statistical analysis, eigenvector calculation, normalized weights, etc.).

**b. Important parameters in preference elicitation methods**

There are some important parameters that have been considered important by several authors to analyse the appropriateness of the preference elicitation methods (Nikou *et al.*, 2015). The parameters were taken from multiples sources which deal with various subjects: consumer research (Nikou *et al.*, 2015, Riabacke *et al.*, 2012), health studies (Cleemput *et al.*, 2014, Marsh *et al.*, 2016). The parameters were mainly adapted from Nikou *et al.* (2015) which compared AHP and CA in their study. Out of those criteria from Nikou *et al.* (2015), one criterion regarding **the method knowledge** is excluded because in most cases respondents are not familiar anyway with the methods. Two publications about choosing appropriate MCDA (Cinelli *et al.*, 2014, Polatidis *et al.*, 2006) were also helpful to define the parameter and rational for evaluation. Table 2 summarizes the description of the important parameters. The extensive description of those parameters is provided below.

1. **Weights typology** : The weights typology depends on whether they generate scaling constant. These constants might represent the rate at which criteria compensate one another (Marsh *et al.*, 2016). When the elicitation task takes into account the performance of other criteria and requires DM to consider to trade-off changes, then the weights most likely to be scaling constants. When the elicitation process includes merely the assessment of the importance of each criterion individually, then the weights might not be scaling constants (Marsh *et al.*, 2016). This trade-off might increase the level of precision since it may represent the true *trade-off* value from the decision maker (Marsh *et al.*, 2016). These qualifications are best satisfied by the swing weighting and decompositional approaches, while direct rating tends to generate more similar weights among criteria (Marsh *et al.*, 2016). On the contrary, importance coefficients reflect the power of the criterion, being independent of other criteria. It is usually expressed with an ordinal meaning (Cinelli *et al.*, 2014). Thus, the weights typology are distinguished between *importance coefficients (no compensation rate is collectively considered)* and *trade-offs* (*compensation is collectively considered)* (Cinelli *et al.*, 2014, Marsh *et al.*, 2016).
2. **Treatment of uncertainty**: This parameter relates to the capability of methods in handling uncertain, imprecise, or missing information/value (Cinelli *et al.*, 2014, Marsh *et al.*, 2016, Riabacke *et al.*, 2012, Nikou *et al.*, 2015). In the diet model, the uncertainty might arise from food price for instance. The price of the food might change seasonally and also differ from one market to another. Intervals might work to deal with imprecise values (Polatidis *et al.*, 2006). Additionally, sometimes respondent leave several questions without answers or being indecisive. This might result in *missing value* issues (Nikou *et al.*, 2015) and *no-choice answer* (Karniouchina *et al.*, 2009).
3. **Robustness**: The method is considered as robust when the addition or deletion of an alternative does not affect the classification or ranking of others. For example, the occurrence of *rank reversal* – reversal in the ranking (Cinelli *et al.*, 2014, Marsh *et al.*, 2016). Rank reversal is a classical problem in AHP which could be due to the inconsistency of the ratio-scale pairwise comparisons (Felli *et al.*, 2008). Nonetheless, more built-in checks for consistency might decrease the chance of rank reversal (Cleemput *et al.*, 2014).
4. **Ease of use** : The term '*ease of use*' is coined by (Cinelli *et al.*, 2014). This term covers the simplicity of the methods based on users (respondents/ DM) and analyst. In other literature, this parameter has different terms, namely '*the complexity of task*' (Nikou *et al.*,

2015) and ‘*feasibility*’ (Riabacke *et al.*, 2012). The elicitation of decision data from the users must be feasible, in terms of the number of required inputs and cognitive load (Riabacke *et al.*, 2012). The complexity of task does not merely depend on the number of attributes and levels, but also the mental or cognitive effort required by respondents (Nikou *et al.*, 2015). The availability of software support to implement the method, manage the information and show the results are highly useful in analysing the data and support the analyst (Cinelli et al., 2014). Moreover, the cognitive load on the DM(s) should also be minimized to eliminate biases and errors that have been acknowledged in behavioural research (Riabacke *et al.*, 2012). Riabacke *et al.* (2012) pointed out that the selection of appropriate elicitation method is a matter of balancing the obtained quality of the elicitation methods with the time and cognitive effort burden on DM(s) when eliciting the required information. Additionally, the analyst’s effort should also be considered especially for the preparation and analyse steps in preference elicitation process.

5. **Sample size** : The potential number of respondents needed also has to be considered (Nikou *et al.*, 2015). If the analyst aims to define diet for a population, then a representative amount of respondents is preferable to increase reliability of the outcome. Some methods need a huge number of respondents to generate reliable and generalizable results (i.e. conjoint analysis, machine learning) (Nikou *et al.*, 2015). If the sample size is too large, then it will be troublesome for the analyst.
6. **Learning dimension** : According to Nikou *et al.* (2015), the structure of the elicitation process is important because it influences the possibility of re-evaluating results if new information becomes available (i.e. alternatives or criteria). This parameter depends on the design of the preference elicitation method. Hence this parameter is called ‘*design*’ in Nikou et al. (2015). However, that terminology might be too general. Thus, another terminology, ‘*learning dimension*’ from Cinelli et al. (2014) was chosen. In the multi-stages method, more steps needs to be repeated in comparison to the single-stage method (Nikou *et al.*, 2015). Hence, it is required to re-run the software and obtain an updated result.
7. **The context of use** : The familiarity of the domain and context of use are two different parameters (Nikou *et al.*, 2015). Nevertheless, familiarity with the domain (i.e. product or service) is closely connected to the context of use (i.e. determine users’ needs or screen product or service) (Nikou *et al.*, 2015). Therefore, these parameters were merged into one under this parameter. To illustrate, in the early stage of new product development, the familiarity of product use is not relevant because there is no existing product in the market yet. At this stage, the context of preference elicitation method is to determine users’ needs. The familiarity with a product might be important in the development stage of a new concept of an already existing product where the context is to screen or evaluate product and service (Nikou *et al.*, 2015). Thus, the knowledge-level of the consumers and the understanding of the context are necessary to be considered.

*Table 2.Short description of important parameters of preference elicitation method*

| Criterion | Description |
|---|---|
| Weights typology | Significance of the weights used to assign importance levels to the criteria |
| Treatment of uncertainty | Capability of handling uncertain, imprecise, and missing information |
| Robustness | Consistency of results after an addition or deletion of alternatives/ criteria |
| Ease of use | Intelligibility of the method, simplicity of the methods based on the respondent (i.e. DM) and analyst perspectives. Availability of software or computer support. |
| Sample size | Requirement of respondents to assure reliability |
| Learning dimension | Possibility of re-evaluating results if new information becomes available (e.g. alternatives or criteria) |
| Context of use | The aim of the preference elicitation used which relates to stage of product/ service development and knowledge-level of respondents |

### 2.2 Diet Modelling

Briend *et al.* (2003) mentioned that mathematical programming (MP) is more efficient and faster to deliver the optimal diet compared to a manual approach (i.e. trial and error). Mathematical Programming (MP) diet model designates the use of mathematical techniques to formulate and optimize diets according to several constraints (Buttriss *et al.*, 2014). Faramitha (2016) reviewed the literature and found out that MPs are the most commonly used approaches. Linear programming is an example of the most basic MP. The previous research of Faramitha (2016) determined the basic components of MP SHARP diet model.

***Decision variables***
Decision variables describe the quantity that would like to be determinedthe as the outcome of the model. In the case of diet model, the decision variables will be the recommended intake of food item $i$ in adefinite period (i.e. daily or weekly basis).
***Objective function***
The objective function contains the selected criteria to evaluate the feasible solutions. The criteria were adapted from optimization aim. In MP diet model, the aim is to create an optimal diet that is sustainable, healthy, affordable, reliable, and preferable. Since SHARP diet has five dimensions, it means that there will be multiple performance indicators that must be evaluated. Thus, it might have a multi-objective function.
***Constraints***
Constraints are the factors that limit the optimization. For example, in a MP diet modelling, an optimal diet is expected to provide sufficient nutrients and energy. Thus, the estimated amount of nutrients that should be improved or should be limited can be included as constraints.
A sensible SHARP diet must also consider preferability. To be more precise, diet recommendations should be desirable and acceptable for consumer targets. Preferability constraints can be helpful to select the preferable foods or limit the food intakes to provide feasible diet recommendations. There are several examples of preferability constraints: 1) creating food groups to simplify diet model; 2) excluding unpopular foods to create acceptable diets; 3) limiting the optimal weight intake of food item $i$ to recommend reasonable amount of food.
***Parameters***
The selected SHARP indicators are the bases to define parameters. For instance, a relevant parameter for affordability dimension is the price of food item.

To give an illustration of MP diet model, we present an example from a study of Ribal *et al.* (2016) which aimed to design 20-day lunch menu for school. In this study, they started by picking 20 starters, 20 main courses, and 7 desserts to provide variety. So, they came up with 2800 lunch menus from those combinations. The **decision variables** are the food items which are divided into a starter, main dish, and dessert. The lunch menus were selected by using goal programming (GP) models. GP model usually has an **objective function** that consists of minimizing the undesirable positive or negative deviations from several goals based on daily average (i.e. nutrient content, budget, calorie content, etc.). The goals can be derived from the **parameters**. In the objective function, each goal has a weight that represents the relative importance of parameter to the decision maker. The weights are called **preference weights**.

### 2.3 Conceptual Review Framework

Figure 8 is the conceptual review framework of present research which depicts the steps to be conducted for this research process. This conceptual review framework combines three main activities that should be done to analyse the obtained preference elicitation methods. The three main activities are classifying, understanding, and evaluating. The 'classifying' activity aimed to identify the obtained preference methods and under which sub-category of decompositional and compositional they belonged to. The 'understanding' activity consisted of elaborating the description of preference elicitation methods with illustrative examples; identifying their main components; and capturing the current trends in their applications. Afterwards, the 'evaluating' activity was carried out to analyse the advantages and disadvantages of the methods and when the methods could be appropriately applied in SHARP diet model.

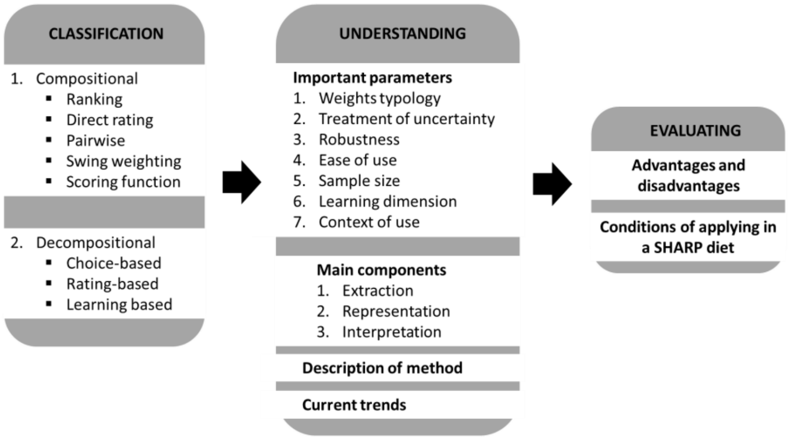


*Figure 8.Conceptual review framework for this present research*

# CHAPTER III. RESEARCH METHODOLOGY

This present research conducted a literature review to identify the typical existing preference elicitation methods and to evaluate when those methods could be appropriately applied in SHARP diet model. There are three key phases that were presented in this chapter: 3.1) identification of relevant research literature; 3.2) data extraction; and 3.3) data analysis.

## 3.1 Identification of relevant research literature

Firstly, the publications were collected by using two search strategies: snowball method and systematic search. Initially, snowball techniques found the relevant articles from the leading articles. From those articles, we could find the keywords that are useful for defining search terms in systematic search methods.

### 3.1.1 Literature search strategies

**SNOWBALL METHOD**

This is used to identify numerous relevant publications based on several leading publications. The process of snowball method is illustrated in the following Figure 9. There are three main steps: (i) identification of leading articles; (ii) tracking relevant articles (*reference tracking* and *cited-by*); and (iii) screening the title, abstract, and keywords. The steps are explained below:

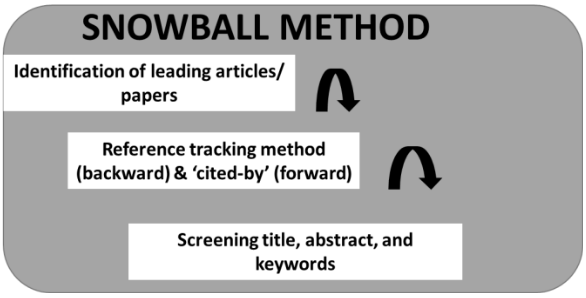


*Figure 9.Snowball method process used in present research*

***Step 1: Identification of leading publications***

The first step is identifying the leading articles. In order to identify the initial set of the leading publications, it is important to specify some criteria to ensure the reliability, relevance, and accuracy of their contents. The criteria are different to the ones already mentioned. These leading publications should be able to capture representative state-of the-art methods of preference elicitation methods in order to provide the foundation for the review. Thus, the leading publications were chosen based on the following criteria:

1. The publications were preferably review papers which provide an overview of cutting-edge information about preference elicitation methods
2. The publications should be recently published (within 5 years)
3. The publications should be highly cited
4. The publications were published in highly ranked journals in the field of operational research, expert, and management science/consumer studies/ health

The publications which were considered as leading publications are presented in Table 3. They represented three application groups, namely MCDA, conjoint analysis, and machine learning. The leading publications for machine learning concentrated on two main methods, collaborative filtering and PageRank. Both methods were selected among the abundant amount of machine learning methods in accordance with supervisors' suggestions. These methods are widely utilized (Elahi *et al.*, 2016, Franceschet, 2011) and relatively quite easy to comprehend compared to other machine learning techniques (i.e. Amazon, Google search).

*Table 3.List of leading publications*

| Authors | Title | Content |
|---|---|---|
| **Multi-Criteria Decision Making** | | |
| **Marsh et al., 2016** | Multiple Criteria Decision Analysis for Health Care Decision Making—An Introduction: Report 2 of the ISPOR MCDA Emerging Good | A structured review of compositional and decompositional elicitation methods |
| **Riabacke et al., 2012** | State-of-the-art prescriptive criteria weight elicitation | A comprehensive review of weight elicitation methods and their classification |
| **Artificial Intelligence, machine learning, and recommender system** | | |
| **Elahi et al., 2016** | A survey of active learning in collaborative filtering recommender systems | A short summary of different recommender system methods and collaborative filtering. |
| **Franceschet, 2011** | PageRank: Standing on the shoulders of Giants | A summary of PageRank development with clear example |
| **Conjoint analysis** | | |
| **Asioli et al., 2016** | Comparison of rating-based and choice-based conjoint analysis models. A case study based on preferences for iced coffee in Norway | Comparison between choice-based and ratings-based conjoint analysis with decent reference list and practical examples |

***Step 2: Reference tracking and cited-by tracking***

After obtaining the leading publications, the snowball technique was carried out to collect more relevant publications. In the *backward procedure*, the reference tracking method was used. This was done by screening the references from the leading publications to obtain earlier relevant publications. The collected results are compiled to the list of publications in Appendix II. In order to check their relevance, the abstract and keywords were screened via Scopus and Web of Science. Scopus and Web of Science were selected as the database because they provide literature from a broad subject area. The tracked publications were mainly the ones being referred by leading publications as sources for certain methods.

The 'cited-by' tracking (*forward procedure*) aimed to find more recent publications for the review as depicted in Figure 9. Since articles of Riabacke *et al.* (2012) and Franceschet (2011) were not very recent, we also used the 'cited-by tracking method to identify more recent articles. However, we could not find a more recent review paper. There was one review paper found, but it reviewed MCDA methods only. While for Franceschet (2011), the other publications were too technical and hard to understand. Hence, we still based on Riabacke et al. (2012) and Franceschet (2011).

Several criteria are specified to find relevant publications from a number of search results while conducting literature research, they were: (1) limited to book, scientific journals, review paper, conference proceeding, report, and trustworthy websites (government or international organization website); (2) written in English; (3) published between 2001 – 2016; (4) related to operational research, expert systems, and management science/consumer studies/ health. Publications which were not related to preference elicitation methods were excluded.

**SYSTEMATIC SEARCH**

A systematic search was an additional search strategy to observe the current trends of preference elicitation method. Generally, systematic search consists of four steps: (i) identifying the key concepts, (ii) formulating search terms, (iii) mixing and matching the search terms and apply those to database tools to find the article results

***Step 1: Identifying the key concepts and keywords***

Initially, we listed the keywords from each identified publication from snowball method and group the similar keywords into key concepts. They were two key concepts, namely elicitation (the type of information that can be derived from preference elicitation) and application (group applications). The selected keywords in terms of elicitation were *preference elicitation method*, *weight elicitation,* *and utility elicitation*. Meanwhile, the other group of keywords consisted of the group applications: *multi criteria decision, machine learning, and conjoint analysis*.

*Table 4.List of key concepts and search terms*

| Key concepts | Elicitation | Application |
|---|---|---|
| **Search terms** | Preference elicitation | Conjoint analysis |
| | Utility elicitation | Machine learning |
| | Weight? elicitation | Multi criteria decision |

**Step 2: Formulating search terms**

For each key concept, we used the keywords as search terms as shown in Table 4. Wildcards symbols such as an asterisk (*) and question mark (?) used to optimize the search result and to make the search terms more efficient. The use of question mark (?) is proposed for only one letter change. For instance, by using search term *weight?*, the database could find the article that contains the word of '*weight*' or '*weights*'.

**Step 3: Combining search terms and applying database tools**

The search terms were joined and matched by using a Boolean operator such as 'AND' and 'OR,' then the search term combinations were applied to selected databases to check its effectiveness in giving relevant literature results. When the number of search results was unrealistic, trial and error step was carried out in defining the most suitable search term combination by adjusting the combination formulation, using more or less specific terms. After that, the inclusion and exclusion criteria were applied in databases to limit the number of relevant publications. The applied screening criteria were:
(1) The publications should be published within 2001 to 2016
(2) The publications were written in English
(3) The publications include a journal article, review paper, and conference paper

Eventually, title, abstract, and keywords of each search result were screened to identify the final relevant publications. The publications should provide real case application of preference elicitation methods. Table 5 indicates the process of refining search terms for Scopus and Web of Science databases. The results showed a tremendous decrease in the number of publications after adjusting the search terms.

*Table 5.Search term combinations in Scopus*

| | | | **Number of results** | |
|---|---|---|---|---|
| **No** | Activities | Search term combination | SCOPUS | Web of Science |
| **1** | Combination of all search terms | TITLE-ABS-KEY ( **"preference elicitation"** OR **"weight? elicitation"** OR **"utility elicitation"** OR **"machine learning"** OR **"multi criteria decision"** OR **"conjoint analysis"** ) | 84,932 | 31,157 |
| **2** | Adjusting the combinations and reducing the search terms | TITLE-ABS-KEY ( **"preference elicitation"** OR **"weight? elicitation"** OR **"utility elicitation"** ) AND TITLE-ABS-KEY ( **"multi criteria decision making"** OR **"conjoint analysis"** OR **"machine learning"** ) | 74 | 55 |
| **3** | Applying exclusion &inclusion criteria | (same as above, but in refined results: specified some criteria) Limit to: published year (2001-2016), languages (English), source type/ document type (article, review & conference paper) | 72 | 52 |
| **4** | Screening the title, abstract, and keywords | | 53 | 39 |

The screening step focused on title, abstract and keywords that should relate to preference elicitation method (i.e. the name of methods, the name of categories, etc.). The collected publications from Scopus and Web of Science were combined, and the similar publications were considered only once. Within one literature, there might be more than one preference elicitation used particularly when the objective was to compare some methods. The publications originated from various fields of applications, including health economic, business/ marketing, and computer/ mathematics.

### 3.2 Data extraction

The objective of data extraction activity was to obtain the important information from relevant publications in relation to preference elicitation methods. Literature review questions were useful to guide the selection of appropriate essential information for each publication. The questions were adapted from conceptual review framework in the previous Chapter 2. Table 6 below presents the list of questions and which sub-question (SQ) they tried to answer.

*Table 6.List of literature review questions*

| No | Questions | |
|---|---|---|
| 1 | What was/were the aim(s) of the research? | |
| 2 | What type of preference elicitation method(s) was/were discussed? | |
| 3 | What was/were the underlying mechanism(s) of the method(s)? (***SQ1: steps of each method***) | |
| 4 | Which category does each method belong to? (***SQ4 : trends of preference elicitation methods***) | |
| 5 | How to apply this in diet model? (description with illustrative examples) (***SQ1: steps of the method***) | |
| | ***The main components (SQ2: Features/ main differences of preference elicitation methods)*** | |
| 5 | Extraction | How is information derived from decision maker's input? |
| 6 | Representation | How do (es) the decision maker(s) represent the decision maker's input? |
| 7 | Interpretation | How does the analyst interpret the data? |
| | ***Important parameters (SQ3: Advantages and disadvantages of each method)*** | |
| 8 | Weights typology | Does the weight reflect trade-off? |
| 9 | Uncertainty treatment | Can this manage uncertain and imprecise information? |
| 10 | Robustness | Is the result dependent on the addition or deletion of alternatives or criteria?<br>Does it have an internal consistency check? |
| 11 | Ease of use | Can the decision maker comprehend the method?<br>Does it require high cognitive effort from decision maker?<br>Does it require high cognitive effort from an analyst? |
| | | Is there software to support the calculation? |
| 12 | Sample size | How many respondents are needed to represent the population? |
| 13 | Learning dimension | Is re-evaluation possible without running it individually? |
| 14 | Context of use | At which stage of development the methods are usually used? |
| | **Conditions of application** (***SQ5: When the methods can be used for SHARP diet model?***) | |
| 15 | SHARP criteria | How many SHARP criteria are considered? |
| 16 | Type of data collection | An interactive or empirical approach is used for data collection? |
| 17 | Aggregation level of consumer | Is the optimization unit based on individual(s) or average-individual(s)? |
| 18 | Dealing with uncertain data | How do the methods deal with uncertain data, deterministic or stochastic? |

### 3.3 Data analysis

This section emphasizes on how the analysis process was carried out for each result section in Chapter 4.

***<u>Classification of the obtained preference elicitation methods</u>***

The collected preference elicitation methods from the publications of snowball method were analysed by using questions 1-4. The goal was to understand the procedure of the methods (SQ 1). After the procedure was known, the methods were classified into either *compositional* or *decompositional* category and also under a sub-category. The definition of those categories and sub-categories could be found in Section 2.1.4. The results were compiled in a table consisting of the *category, the sub-category, the name of the method, brief summary procedure, and the source of publications*.

***Trends in preference elicitation methods***

The relevant publications from systematic search were read to determine what preference elicitation method they applied and in which category they belonged to (question 4 in Table 6). In an Excel table, the methods were ordered based on the publication year and recorded as either *compositional* or *decompositional*. The total number of their implementation in literature was counted for each year and accumulated over the years (2001-2016) to capture a trend. A line chart was created to display the trend of compositional and decompositional methods' implementation over time (from 2001 to 2016). Additionally, the methods were also organised based on the group of applications: conjoint analysis, machine learning, or MCDA. The data was recorded in an Excel table and exported to a pie chat. The result (in %) indicated the mostly used group of applications from 2001-2016.

***Description of the preference elicitation methods***

To provide a better overview regarding the methods, we provided illustrative examples for all obtained methods in diet model context. These examples were adapted from the cases in relevant publications. The SHARP diet criteria to be incorporated into the examples were chosen based on Table 1. For decompositional methods, the criteria levels were also determined. The examples were presented as step-by-step procedure and complemented with a brief theoretical introduction. This was a complementary to answer SQ 1.

***Main components of the methods***

In order to understand the features or main differences of each method (SQ 2), the methods were evaluated according to question 5-7 in Table 6. The definition of each main component was elaborated in Section 2.1.6 *Characteristics of Elicitation Methods*. We used the articles of Nikou *et al.* (2015), Marsh *et al.* (2016) and Riabacke *et al.* (2012) as guidelines to define the main components of the methods. They provided tables containing the main components of several preference elicitation methods. When the method was not on the list, we used the step-by-step procedure as a benchmark to derive the main components of the method. The results were then presented in a table comprising of *the methods and the main components (i.e. extraction, representation, and interpretation*).

***Advantages and disadvantages of the methods***

To comprehend the advantages and disadvantages of each method (SQ 3), we had to evaluate the important parameters based on questions 8 to 14 in Table 6.Table 7 below provides the guideline. As the advantage and disadvantage were not relevant issues in context use, it was not included in the rational for evaluation.

*Table 7.Rational evaluation of important parameters (adapted from Polatidis et al. (2006) and Nikou et al. (2015))*

| Criterion | Rational for evaluation | |
|---|---|---|
| | **Advantage** | **Disadvantage** |
| Weights typology | Weights reflect only the importance coefficients, so compensation is not indicated. | Weights reflect the trade-offs which indicate compensation and equal proportion. |
| Uncertainty treatment | Uncertain and imprecise information cannot be managed. | Uncertain and imprecise information can be managed. There is an assessment to perform consistency check (i.e. test and retest) |
| Robustness | Results are dependent on addition or deletion of alternatives/ criteria (i.e. *rank reversal*). No assessment to perform an internal consistency check. | Results are independent of new alternatives/criteria or deletion of existing ones. There is a constant internal consistency check. |
| Ease of use | The method is perceived as a black-box from the decision make, it is highly demanding in terms of cognitive efforts for the decision maker, and it requires a great effort for analysts to prepare and analyse. | Intelligibility of the method(s) is very simple, the decision maker and analysts are comfortable with the preferences elicitation process |
| | Limited availability of software and poor graphical representation | software available and wide range of graphical potentials that improves the communication with stakeholders |
| Sample size | A large sample size of respondents to represent population | A small sample size of respondents to represent population |
| Learning dimension | No re-evaluation is possible and new software runs need to be performed and independently compared with the previous ones | Assessments can be run with new alternatives and compared simultaneously |

***When the methods can be applied in diet modelling***

To understand under which conditions the methods can be applied in SHARP diet modelling (RQ 5), the methods were evaluated according to four conditions which were adapted from Faramitha (2016). The list of questions was presented in Table 6 (question 15 -18), but the extensive explanation is provided below.

***- The identification of indicators for SHARP diet***

The identification of indicators for SHARP diet is a crucial initial step because each SHARP dimension might be represented by multiple indicators (so-called *criteria*). For example, the sustainability dimension can be indicated by GHGEs emission, water use, waste production, and land use. Most of the time, the analyst would choose the most representative indicator (criteria) for every dimension. The main concern is that the more criteria to be considered, the more cognitive burden is imposed to respondents. So, the goal was to know which method(s) was (were) suitable when the all of the SHARP criteria were going to be considered and when only some of them were considered.

***-The type of data collection***

According to the method of collecting data, the elicitation can be categorized into two main approaches, namely *interactive and non-interactive approaches* (Chankong and Haimes, 2008). An interactive approach implies that there is an active progressive interaction between the DM and analyst in several stages. So, it demands a high involvement of DM. Meanwhile, a non-interactive is related to an empirical approach which involves a data set of observations to recover the preferences of DMs. The empirical approach requires abundant of data. Hence, the objective was to know which method(s) was (were) suitable for interactive and empirical approaches.

***-The aggregation level of consumer***

The analyst needs to specify the optimization units that characterize the respective consumer group(s). It will relate to the group aggregation level. The main question would be whether the MP diet modelling should optimize a recommended diet solution for based on the preference of **average-individuals** within the consumer group or only preference of **an individual (or several individuals)** which represent(s) the target consumer. For the average-individuals approach, the outcome is obtained from the preference of the dietary data of individuals within the target group. For the latter approach, the MP diet model only optimizes a diet recommendation for an individual (or some individuals) as the reference. For a huge number of people, an interactive method might be burdensome to apply since it requires the high involvement of DM. It will be easier to use an interactive method for an individual (or some individuals) to get more comprehensive and reliable information. Thus, the question is which method(s) was (were) more applicable for average-individuals and which method(s) was (were) for the indvidual(s).

-**Dealing with uncertain data**

The concern might arise when collecting input for criteria due to the data uncertainty. For example, the price of food might differ depending on where the food is sold. The analyst should decide whether the variation will be taken into account or not because some target groups might be price sensitive. So, it might affect their preference towards certain diet. In addition, there are numerous types of uncertainty that can be observed, such as GHGE data. The analyst could hardly find GHGE data for all food or drink items. Thus, an estimation should sometimes be made. According to Faramitha (2016), there are two main approaches in dealing with that issues, *deterministic (no uncertainty is considered) or stochastics (uncertainty is considered) approach*. In deterministic approach, there are a few strategies that could be used: referring to the most important wholesaler (Ribal *et al.*, 2016) or food price database such as from USDA (Metzgar *et al.*, 2011). For stochastic approach, the price can be estimated through deflating the present food prices with Consumer Price Index (CPI) (Håkansson, 2015). In order to cope with that, the preference elicitation methods can have various strategies. One of them is by introducing an interval or a category scale to accommodate a range of values. Thus, the goal was to understand which methods that used deterministic and stochastic approach respectively.

# CHAPTER IV.RESULTS

This chapter presents the outcomes of the literature review. This section is started with presenting the identified publications. Then, it is followed by the classification of reviewed methods, the current trends in preference elicitation methods, the description of the methods, their main components, advantages and disadvantages, as well as under which conditions the methods can be applied in SHARP diet model.

## *4.1 Identified relevant publications*

Table 8 presents the results of identified publications for both snowball method and systematic literature review. From the snowball method, there were 42 relevant publications selected from 5 leading publications. The list of publications is summarized in Appendix II. During the screening, a number of publications were excluded because they provided the same contents, so they did not add additional information. For example, most of the tracked publications from Asioli *et al.* (2016) delivered the similar contents about choice-based and rating-based conjoint analysis. Thus, in order to prevent redundant information on conjoint analysis itself, only several recent articles were thoroughly reviewed.

From the systematic search, there were 58 relevant publications found. A lot of publications were eliminated because they were duplicates (22 papers). The list of publications for systematic search can be accessed in Appendix I. About 12 publications were excluded after skim reading because they did not provide an application of the method in a real case situation. The publications that merely discussed the advantages and disadvantages of the methods were not included for systematic search but were kept for analysis, such as (Louviere *et al.*, 2010) and Lloyd (2003). The article of Louviere focused on the theoretical comparison between the DCE and Conjoint Analysis, while Lloyd (2003) revealed the consequences of using different preference elicitation methods.

*Table 8.The results of snowball method and systematic literature research*

| | **Snowball method** | | **Systematic search** | |
|---|---|---|---|---|
| | Reference tracking | Cited-by | SCOPUS | Web of Science |
| *1. Screening the title, abstract, and keywords* | 40 | 2 | 53 | 39 |
| *2. Combining all publications (eliminating duplicate records)* | 40 | 2 | 70 | |
| *3. Skim-reading the full text publications* | 42 | | 58 | |

## *4.2 Classification of obtained preference elicitation methods*

As mentioned in Section 2.1.4, there are two main categories of preference elicitation methods, namely compositional and decompositional. There are 5 sub-categories under compositional category (*direct rating, ranking, pairwise comparison, SWING weighting, scoring function*), and 3 sub-groups under decompositional group (*rating-based, choice-based, and learning based*). These sub-categories represent how the elicitation processes are conducted. Table 9 and 10 present the classification of reviewed methods together with their procedures and source of publications.

In total, there are 9 methods under compositional group and 6 methods under decompositional group. If the method evaluates criteria first, then it belongs to the *compositional group*. According to the definition of compositional and decompositional (Section 2.1.4), if the method evaluates the utility of alternative(s) first, then it belongs to the *decompositional group*. Based on the observation, there could be several variations of the method under a sub-category. In a compositional category, the *direct rating* has the most variations of the method under its sub-category. The variations include assigning rate directly (i.e. scales), distributing points among criteria (point allocation), combining ranking and rating methods (i.e. SMART), and using interval value (i.e. interval SMART). In several sub-categories, there is only one method found (i.e. SWING weighting and scoring function).

*Table 9.The classification of compositional method*

| Compositional | | | |
|---|---|---|---|
| **Group** | Method | Brief summary of procedure | Sources |
| **Direct rating** | Scales | Rate each criterion on 0-100 scale or 1 -5 scale, etc. | Riabacke et al., 2012; Goetghebeur et al., 2012; Bottomley and Doyle, 2001; Hein et al., 2008; Pöyhönen and Hämäläinen, 2001 |
| | Point Allocation | Distribute 100 points among criteria | Riabacke et al., 2012; Kroese et al., 2010; Bottomley and Doyle, 2001; Pöyhönen and Hämäläinen, 2001 |
| | Simple Multi-Attribute Rating Technique (SMART) | Rank the criteria and assign the least important criterion by 10 points. Rate the remaining criteria relative to the least important one. | Riabacke et al., 2012; Pöyhönen and Hämäläinen, 2001 |
| | Interval SMART | Rate the reference criterion with a fixed point, and then provide interval values to indicate other criteria relative to the reference criterion. | Mustajoki et al., 2005; Riabacke et al., 2012 |
| **Ranking** | Direct ranking | Ordinal statements of criteria importance (rank the criteria from most important to the least one. | Marsh et al., 2016; Cleemput et al., 2014; Danielson and Ekenberg, 2016 |
| | SMART Exploiting Ranks (SMARTER) | Rank the criteria and then convert such rankings to numerical weights. | Riabacke et al., 2012; Pöyhönen and Hämäläinen, 2001 |
| **Pairwise comparison** | Analytic Hierarchy Process (AHP) | Compare the each criterion in pairs based their relative importance towards each other by using 1-9 ratio scale. | Marsh et al., 2016; Cleemput et al., 2014; Nikou et al., 2015; Dolan et al., 2005; Pöyhönen and Hämäläinen, 2001 |
| | Measuring Attractiveness by a Categorical-Based Evaluation Technique (MACBETH) | Compare the each criterion in pairs based their relative importance towards each other by using semantic categories. | Marsh et al., 2016; Oliviera et al., 2012; Pinhero et al., 2008; e Costa et al., 2012 |
| **SWING weighting** | SMART Swing (SMARTS) | Rank the criteria and consider them at the worst level. Swing each criterion from the worst to the best level and assign 100 to the most important criterion. Rate the other criteria relative to the most important criterion. | Marsh et al., 2016; European Medicines Agency, 2011; Felli et al., 2009; Pöyhönen and Hämäläinen, 2001 |
| **Scoring function** | Bisection | Define the importance of a criterion using the value function (0-100) by identifying the mid-point. | Marsh et al., 2016; Tervonen, 2015; Belton and Stewart, 2001 |

Based on the reviewed methods within the decompositional category, the sub-category of *choice-based* has more variations than others. They vary on choosing only the best alternative (CBCA), the best and the worst (BWS), and the best alternative if only two criteria are given (PAPRIKA). The variations might aim to minimize cognitive burden imposed on the DM. For examples, comparing alternatives which differ on two criteria might be easier than which differ on multiple criteria.
To sum up, the methods can be classified into the either decompositional (direct rating, ranking, pairwise comparison, SWING weighting) and compositional (choice-based, rating-based, or learning-based). There could be several variations of the method under one sub-category. The modifications might be simple, but it might have an influence on the cognitive process of the DM.

*Table 10. The classification of decompositional methods*

| Decompositional | | | |
|---|---|---|---|
| **Group** | Method | Brief summary of procedure | Source |
| **Choice-based** | PAPRIKA | Choose one alternative, given the performance of each on two criteria. | Marsh et al., 2016; Hansen et al., 2012; Golan & Hansen, 2012; Johnson et al., 2014; French et al., 2015; Cleemput et al., 2014 |
| | Best worst scaling (BWS) | Choose the best and the worst alternatives/criterion's levels from three or more choices | Marsh et al., 2016; Cleemput et al., 2014; Al Janabi et al., 2011; Swancutt et al., 2008 |
| | Choice-based Conjoint Analysis (CBCA)/ Discrete Choice Experiment (DCE) | Choose one alternative, given the performance of each on all the criteria | Marsh et al., 2016; Asioli et al., 2013; Almli et al., 2015; Wezemeel et al., 2014; Saito, 2012; Karniouchina et al., 2008; Baltussen et al., 2007; Marsh et al., 2012; Defechereux et al., 2012 |
| **Rating-based** | Rating-based conjoint analysis (RBCA) | Rate the alternatives, given the performance of each on all the criteria. | Marsh et al., 2016; Asioli et al., 2013; Almli et al., 2015; Gracia and de-Magistris, 2013; Annunziata & Vecchio, 2012; Karniouchina et al., 2008 |
| **Learning-based** | Collaborative filtering | Recommend items which are highly rated by similar users. Users are similar if they co-rate the items similarly. | Elahi et al., 2016; Koren & Bell, 2011; Desrosiers & Karypis, 2011; Adomavicius & Tuzhilin, 2006 |
| | PageRank | Rank the alternatives according to their importance derived from buying frequency of each alternative | Franceshet, 2011; Grover, 2012 |

## 4.3 Trends in preference elicitation methods

Figure 10 presents trends in preference elicitation methods in the literature based on systematic search. It could be seen that the usage of decompositional methods in literature dominated compositional methods from 2001 to 2016. Decompositional methods could be more attractive because they are usually based on real purchase behaviour or at least trying to mimic the actual choice behaviour of respondents by assessing the whole product profiles (i.e. conjoint analysis). This might decrease the bias and increase the validity of the outcome.

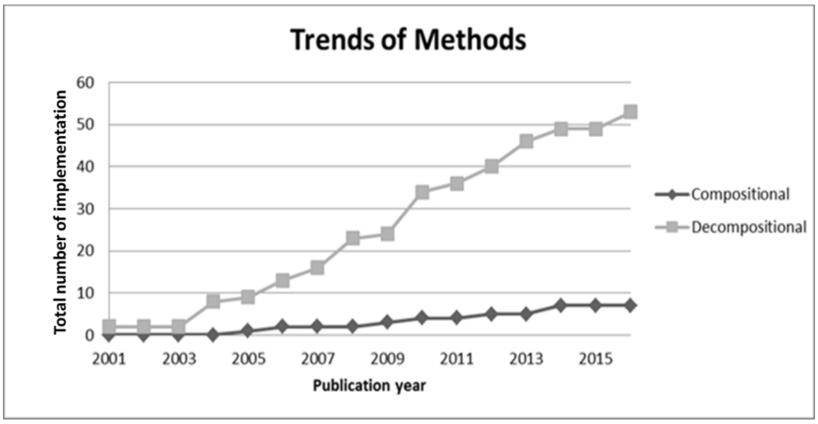


*Figure 10.The trends of preference elicitation methods*

Figure 11 illustrates the recapitulation of preference elicitation methods based on group of applications. The most frequently used is conjoint analysis methods with 65% (i.e. DCE/CBCA, RBCA, BWS, ACA), followed by machine learning methods (i.e. collaborative filtering, recommender system, etc.), and MCDA (i.e. AHP, direct rating). Conjoint analysis approaches might be popular because they could mimic the actual behaviour of consumer yet not require too large data set as machine learning methods. The analysis of data is also relatively simpler compared to machine learning techniques because the supporting software is readily available. For machine learning technique, the analyst might need to create a specific algorithm that is appropriate for this diet case.

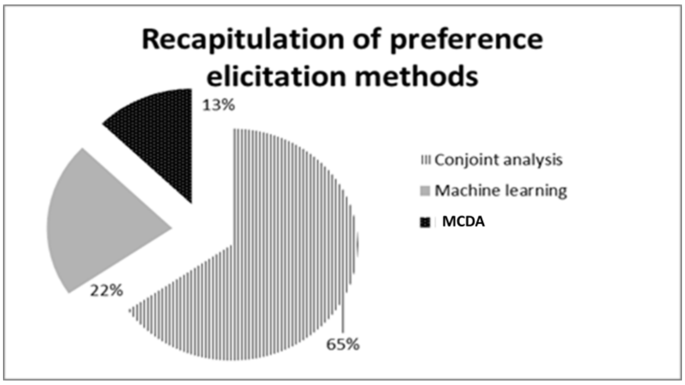


*Figure 11.Recapitulation of preference elicitation method from systematic search*

Several methods were observed in systematic search, but not captured in reviewed methods, such as Time Trade-Off (TTO) and Adaptive Conjoint Analysis (ACA). TTO is usually used in health economic. TTO is often applied to value health states by asking the individuals to trade-off years of perfect health (TTO) (Ali and Ronaldson, 2012). This method resembles bisection method in terms of eliciting trade off. The difference is TTO requires respondents to define how many years in a healthy state would be equivalent to x years in a poorer state of health. The utility or value for each outcome will be calculated from the comparison of those time units. Since this method evaluates alternative(s), they might be categorized as decompositional method.

ACA is an extension of CBCA. The aim is to provide more engaging elicitation process than conventional CBCA and improve the prediction of actual preferences (Cunningham *et al.*, 2010). This method is more appropriate for larger of amount criteria than CBC. Since ACA could capture more comprehensive data than the traditional CBCA, the outcome is more stable and might need smaller sample size. Yet, it also takes more time to finish the task compared to CBCA (Cunningham *et al.*, 2010). ACA is also considered as decompositional method.

In sum, there were two main findings from systematic search. Firstly, decompositional methods dominated compositional method in terms of frequency of application based on systematic search. Secondly, the conjoint analysis methods were the most frequently used in comparison to machine learning and MCDA.

## *4.4 Description of preference elicitation methods*

This section provides a brief explanation of obtained methods, accompanied with illustrative diet-related examples. The possible criteria for the illustrative examples must be determined. The factors were adapted from the elements and indicators of SHARP diet found by Faramitha (2016). Table 11 presents the basic criteria of each SHARP dimension, their corresponding indicators as well as a range of value. This information will be integrated and adjusted for the illustrative examples for each method. This section is divided into two main parts, namely description for *compositional* and *decompositional* methods.

*Table 11. The indicators of SHARP diet criteria for illustrative examples*

| ***Dimension*** | ***Sustainability*** | ***Health*** | ***Affordability*** | ***Preferability*** |
|---|---|---|---|---|
| ***Criteria*** | ***Environment*** | ***Health*** | ***Cost*** | ***Food intake*** |
| **Performance Indicator** | Reduction of GHGEs emission | Fulfilment of nutritional requirement | Food price | Food intake limit |
| **Units** | % | % | euros | gram/day |

Direct rating
With the direct rating technique, the DM assigns a score to denote the importance of each criterion.

- **Scales**

Scales indicate the importance of alternatives on each criterion on a scale (Marsh *et al.*, 2016). One of the scales that can be used is *Visual Analogue Scale (VAS).* There are numerous variations of this approach (i.e. the length of the line, labels for the ends of the line, presence or absence of scale marks on the line, and presence or absence of numbers on the scale marks (Torrance *et al.*, 2001). Otherwise, the analyst can simply ask to give rate based on a particular scale (i.e. 0 to 5) (Bottomley *et al.*, 2000). This method allows the decision maker to alter the importance of one criterion without adjusting the weight of another. Thus, the respondents do not have to make trade -off among criteria (Yusop, 2015). This method generally has a tendency to generate similar weights among criteria (Marsh *et al.*, 2016) because the respondents are reluctant to use extreme scale (Torrance et al., 2001). This behaviour is known as ***end aversion*** and affecting the final scores (Torrance et al., 2001).

The best alternative is the one with the highest multi attribute value (MAV). The utility model used is a *weighted additive* (Equation 1). Each alternative $i$ is measured on $M$ (*number of criteria*) , the person gives ratings of the importance of criteria ($j$) and values of $xij$ for *j*-th criteria of the *i*-th alternative (Bottomley *et al.*, 2000).

$$MAV_i = \sum_{j=1}^{M} w_j x_{ij} \qquad \forall\ i \in \{1, \dots, n\} \tag{1}$$

where, $wj$ = weights of criteria $j$; $xij$= values for *j*-th criteria of the *i*-th alternative; $MAVi$ = total multi-attribute value of alternative $i$

**Example**
The example given is using a 5-point scale, with 1 representing the least and 5 the most important criteria.
(1) The respondents are asked to rate criteria on a scale of 1 to 5 (considered as *weights*). Table 12 provides the example of results and calculation.

*Table 12.The example of scales calculation*

| CRITERIA | Respondents | | | | Calculation | |
|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **Mean weights** | **Normalized mean weights** |
| 1. *(% of GHGEs reduction)* | 2 | 1 | 2 | 2 | 1.75 | 0.14 |
| 2. *(Nutritional Content)* | 3 | 2 | 3 | 2 | 2.50 | 0.20 |
| 3. *(Price)* | 4 | 5 | 5 | 5 | 4.75 | 0.38 |
| 4. (*Food intake*) | 4 | 3 | 4 | 3 | 3.50 | 0.28 |
| | | | | Total weights | 12.5 | 1.00 |

(2) The weights are then normalized to sum to one as depicted in Table 12 and useful to calculate the MAV for the alternative $i$.

$$MAV_i = (0.14 * Criterion\ 1) + (0.20 * Criterion\ 2) + (0.38 * Criterion\ 3) + (0.28 * Criterion\ 4)$$

- **Point Allocation**

Similar to scales, point allocation is also based on MAV. The only difference is that commonly the respondent (DM) is asked to assign numbers to describe the weights of each criterion directly (Riabacke *et al.*, 2012). This is a very simple method, but the resulted weights are not very precise because people tend to give nearly 50% more weight to their most important attribute compared to those using direct (Bottomley and Doyle, 2001). Moreover, this method becomes more difficult once the number of criteria increases to 6 or more because the respondents have to keep track on the distributed points (Yusop, 2015).

**Example**

(1) Respondents distribute 100 points among the criteria. The points for all respondents are then computed to acquire the mean weights in Table 13.

*Table 13.The example of Point Allocation case and weights calculation*

| CRITERIA | Respondents | | | | Calculation | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | Mean weights | Normalized mean weights |
| 1. *(% of GHGEs reduction)* | 15 | 9 | 14 | 17 | 13.75 | 0.138 |
| 2. *(Nutritional Content)* | 23 | 18 | 21 | 16 | 19.50 | 0.195 |
| 3. *(Price)* | 32 | 45 | 36 | 42 | 38.75 | 0.387 |
| 4. (*Food intake*) | 30 | 28 | 29 | 25 | 28.00 | 0.280 |
| | | | | Total | 100 | 1.000 |

(2) The weights are then normalized to sum to one as depicted in Table 13 above. The utility function used is also *weighted additive*. The MAV for alternative $i$ will be:

$$MAV_i = (0.138 * Criterion\ 1) + (0.195 * Criterion\ 2) + (0.387 * Criterion\ 3) + (0.280 * Criterion\ 4)$$

- **Simple Multi-Attribute Rating Technique (SMART)**

This method has the similar calculation steps as the previous methods. The major difference is that respondents have to evaluate the relative importance of a criterion compared to the reference criterion, which is the *least important criterion* (Riabacke *et al.*, 2012).

**Example** (adapted from Riabacke *et al.* (2012))

(1) The respondent is asked to choose the least important criterion (out of 4 criteria) and assigned a weight of 10.

(2) The respondent rates of all other are criteria multiples of 10, relative to the least important one. For example, in our case, the ***GHGEs reduction*** is the least important. Then the other three criteria will be judged compared to the importance of ***GHGEs reduction*** as multiples of 10. The resulting weights are then normalized to sum to one. Table 14 provides an example of calculation.

*Table 14.The example of SMART and weights calculation*

| CRITERIA | Respondents | | | | Calculation | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | Mean | Normalized mean weights |
| 1. *% GHGEs reduction* | 10 | 10 | 10 | 10 | 10.00 | 0.113 |
| 2. *Nutritional Content* | 15 | 20 | 15 | 20 | 17.50 | 0.197 |
| 3. *Price* | 20 | 50 | 25 | 50 | 36.25 | 0.408 |
| 4. *Food intake* | 20 | 30 | 20 | 30 | 25.00 | 0.282 |
| | | | | Total | 86.25 | 1.000 |

(3) The weights are then normalized to sum to one as depicted in Table 14 above. The value for alternative $i$ will be:

$$MAV_i = (0.113 * Criterion\ 1) + (0.197 * Criterion\ 2) + (0.408 * Criterion\ 3) + (0.282 * Criterion\ 4)$$

- ***Interval SMART***

This method employs interval judgments to represent imprecision during extraction instead of point estimates (Riabacke *et al.*, 2012). Interval SMART is an extension of SMART method. With Interval SMART, the reference criterion is not necessarily the most or least important one; any criterion can be the reference criterion (Mustajoki *et al.*, 2005b). Moreover, respondent can use interval judgements on the *weight ratio questions*. The term *weight ratio* refers to the practice of assigning any number of points to the criterion as long as it is relative to the reference criterion (Mustajoki *et al.*, 2005b).

**Example** (adapted from Mustajoki *et al.* (2005b))

(1) Respondent needs to choose one reference criterion which is easily measurable. In our case, this could be the *price*. For this example, the **price** will be *criterion 1*, **GHGE reduction** will be *criterion 2,* and **nutrition** will be *criterion 3.*

(2) In practice, respondent can assign any number of points to the reference criterion, as long as the points assigned to the other criteria are relative to this reference criterion.

(3) Respondents may reply with intervals to the weight ratio questions. So, the reference criterion may have a fixed number of points, but the other criteria might be given an interval to represent the imprecision in the judgement. The equation for intervals is shown in Equation 2.

(4) The analyst can determine the feasible region of the weight (S) by using Equation 2 and weight normalization constraint (Equation 3). The lower bound (Equation 4) and upper bound (Equation 5) can be solved with minimization and maximization problems in linear programming as shown below. Lower bound for the overall value of alternative $x$ $\left(\underline{v}(x)\right)$, is elicited as its minimum, by allowing the weights and the lower bound of criteria values ($\underline{v}_j(x_j)$), to vary within the given constraint, that is in Equation 4.

| Lower bound | | Upper bound | |
|---|---|---|---|
| $\underline{v}(x) = \min\limits_{w \in S} \sum_{j=1}^{n} w_j \underline{v}_j(x_j)$ | (4) | $\overline{v}(x) = \max\limits_{w \in S} \sum_{j=1}^{n} w_j \overline{v}_j(x_j)$ | (5) |
| Subject to | | **Subject to** | |
| $\frac{ref}{max_j} \leq \frac{w_{ref}}{w_j} \leq \frac{ref}{min_j} \quad \forall\ j \in \{1, \dots, n\}$ | (2) | $\frac{ref}{max_j} \leq \frac{w_{ref}}{w_j} \leq \frac{ref}{min_j} \quad \forall\ j \in \{1, \dots, n\}$ | (2) |
| $\sum_{j=1}^{n} w_j = 1$ | (3) | $\sum_{j=1}^{n} w_j = 1$ | (3) |
| where, $\underline{v}_j(x_j)$ = lower bound for $v_j(x_j)$; $\overline{v}_j(x_j)$ = upper bound for $v_j((x_j)$, $v_j(x_j)$ = the value/ score of criteria $j$ in alternative $x$; $w = (w_1, \dots, w_n) \in S$; $ref$ = points given to the reference criterion; $max_j$ / $min_j$ = maximum/ minimum number of points given to other non-reference criteria $j$ | | | |

An example of a case is taken from Mustajoki *et al.* (2005b) *if criterion 1: 1, criterion 2 : interval from 0.5 to 2.0 points, criterion 3 : interval from 1.0 to 3.0 points. The weight ratio constraints (from Equation 2) are:* $^{w_1}/_{w_2}$ = *[1.0/2.0, 1.0/0.5] = [1/2,2] and* $^{w_1}/_{w_3}$ = *[1.0/3.0, 1.0/1.0] =[1/3, 1].*

*These values will define the feasible region of the weight (S) (Figure 12). To avoid complexity, we assume that the lower and upper bounds of each rating interval are similar. Additionally, set these for alternative A and B as* = $\underline{v}_1(A) = \overline{v}_1 A = 0.0, \underline{v}_1(B) = \overline{v}_1 B = 1.0, \underline{v}_2(A) = \overline{v}_2 A = 1.0,, \underline{v}_2(B) = \overline{v}_2 B = 0.8, \underline{v}_3(A) = \overline{v}_3 A = 1.0,\ \underline{v}_3(B) = \overline{v}_3 B = 0.0.$

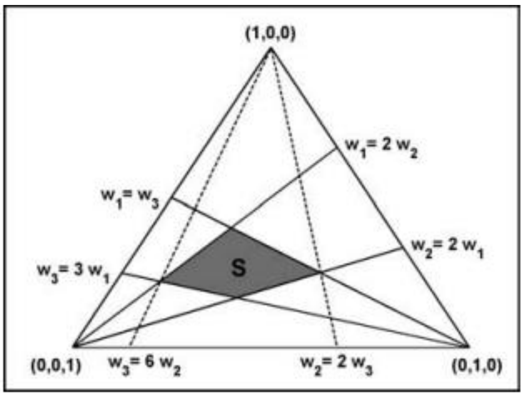

*Figure 12. The feasible region of the weight (S)*

(5) The analyst determines the dominance relations between the alternatives by employing the alternatives' value intervals. Specialized software (i.e. WINPRE (Mustajoki *et al.*, 2005b)) is used to compute these values in order to obtain the overall values intervals for both alternatives. Since there is an upper and lower bound problem, each of the alternatives will have interval values as well. If the value of *alternative A* is bigger than the value of *alternative B* for every possible

combination of the weights, then **alternative *A*** is preferable than **alternative *B***. *In the example case, the overall values intervals for **alternative A** is [0.60, 0.85] and **alternative B** is [0.31, 0.65] (Mustajoki et al., 2005b). Thus, alternative A is the best alternative.*

If no dominating alternative is found, the respondents can provide narrower intervals. When an alternative begins to dominate all the other alternatives, the respondents should decide whether the conditions leading to the narrower intervals are acceptable (Mustajoki *et al.*, 2005a). Empty feasible region indicates inconsistency in the DM's preference assessments. In such a case the DM is requested to rethink his/her preferences.

Ranking

Rank all criteria from the most important to the least important (ordinal statement of importance) (Riabacke *et al.*, 2012).

- **Direct ranking**

Respondents are typically more at ease to provide rankings than precise numbers (Danielson and Ekenberg, 2016). Basically, the criteria are ranked from most important to the least important accordingly. Subsequently, *surrogate weights* can then be derived from the rankings. *Surrogate weights* denote the most representative numerical interpretation of the respondents' ordinal preferences (Danielson and Ekenberg, 2016). The major challenge is to assign surrogate weights while maintaining the 'accuracy' of the weights. For that reason, there are several approaches available to process surrogate weights (so-called *rank-based weighting methods*), namely Rank Sum (RS), Rank Reciprocal (RR), Rank Order Centroid (ROC) (Danielson and Ekenberg, 2016), and Rank Exponent (RE) (Yusop, 2015) as shown in Table 15. RS, RR, and ROC are the most common methods (Danielson and Ekenberg, 2016). The RS reflects the weights directly from the rank order (Equation 6). Meanwhile, RR is constructed from the reciprocals (inverted numbers) of the rank order for each item ranked (Equation 7). RE has several interesting properties. For $p = 0$ results in equal weights for all criteria; while for $p = 1$, the criteria obtain similar weights as RS. In general, the increase of exponent $p$ will give more dominance to the most important criterion (Malczewski, 1999). Table 16 presents how these methods result in different weights for each criterion. The accuracy of the weights from each method is commonly benchmarked against the 'true' weights from respondents. In most recent studies, ROC method is found to outperform RS and RR in preserving the accuracy of predicting DM's actual preference weights (Danielson and Ekenberg, 2016, Ahn, 2011, Riabacke *et al.*, 2012). Thus, ROC is usually recommended to process the surrogate weights.

*Table 15.Table of surrogate weights (adapted from (Danielson and Ekenberg, 2016))*

| | | | |
|---|---|---|---|
| Rank Sum (RS) | $\boldsymbol{Nw_i} = \dfrac{\boldsymbol{N - r_i + 1}}{\sum_{\boldsymbol{j=1}}^{\boldsymbol{N}}(\boldsymbol{N - r_j + 1})}$ | $\forall\ \boldsymbol{i} \in \{\boldsymbol{1}, \dots, \boldsymbol{n}\}$ | (6) |
| Rank Reciprocal (RR) | $Nw_i = \dfrac{1/r_i}{\sum_{j=1}^{N}\left(1/r_j\right)}$ | $\forall\ i \in \{1, \dots, n\}$ | (7) |
| Rank Exponent (RE) | $Nw_i = \dfrac{(N - r_i + 1)^p}{\sum_{j=1}^{N}(N - r_j + 1)^p}$ | $\forall\ i \in \{1, \dots, n\}$ | (8) |
| Rank Order Centroid (ROC) | $Nw_i = \dfrac{1}{n}\sum_{j=i}^{n}\dfrac{1}{r_j}$ | $\forall\ i \in \{1, \dots, n\}$ | (9) |

*where, $\boldsymbol{Nw_i}$ = normalized weight for the criterion $\boldsymbol{i}$; $\boldsymbol{r_i}$ = the rank position of the criterion $\boldsymbol{i}$; $\boldsymbol{N}$= number of criteria under consideration ($\boldsymbol{1}, \boldsymbol{2}, \boldsymbol{3}, \boldsymbol{4}$), and $\boldsymbol{r_j}$ = the rank position of the other criteria*

**Example :**

According to our case, the respondent indicates the rank by giving the number to the criteria (1=*most important* to 4 = *least important*). Table 15 presents the ranks and their calculations with different methods. It could be seen that the calculated surrogate weights (denoted by the normalized weights) vary among different rank-based weighting methods. In comparison to other methods, ROC showed to generate higher weights for the most important criterion.

*Table 16.The example of surrogate weights*

| | | Rank sum | | Rank reciprocal (p=2) | | Rank exponent | | Rank order centroid | |
|---|---|---|---|---|---|---|---|---|---|
| **Criteria** | $r_i$ | $w_i$ | $Nw_i$ | $w_i$ | $Nw_i$ | $w_i$ | $Nw_i$ | $w_i$ | $Nw_i$ |
| *1. GHGEs reduction* | 4 | 1 | 0.1 | 0.250 | 0.120 | 1 | 0.033 | 0+0+0+1/4 = 0.25 | 0.0625 |
| *2. Nutritional Content* | 3 | 2 | 0.2 | 0.333 | 0.160 | 4 | 0.133 | 0+0+1/3+1/4=0.58 | 0.1458 |
| *3. Price* | 1 | 4 | 0.4 | 1.000 | 0.480 | 16 | 0.533 | 1+1/2+1/3+1/4=2.08 | 0.5208 |
| *4. Food intake* | 2 | 3 | 0.3 | 0.500 | 0.240 | 9 | 0.300 | 0+1/2+1/3+1/4=1.08 | 0.2708 |
| Total | | 10 | 1.0 | 2.083 | 1.000 | 30 | 1.000 | 4.00 | 1.0000 |

where : $\boldsymbol{r_i}$ = rating of criterion $\boldsymbol{i}$; $\boldsymbol{w_i}$ = weight of criterion $\boldsymbol{i}$; $\boldsymbol{Nw_i}$ = normalized weight of criterion $\boldsymbol{i}$

In addition to surrogate weights, ELICIT method is also useful to process ranking (Diaby *et al.*, 2016). It is a relatively new method which starts with a rank ordering of the criteria. Then, the representation of data makes use of PCA (Principal Component Analysis) as an aggregation tool for the group criteria ranking. PCA aims to find the best linear combination of variables (i.e. best weighting set) that explains the largest part of the data dispersion while maintaining ordinal consistency. The estimation of criteria weight is continued by Monte Carlo simulation. This simulation consists of substituting point estimates of parameters with inherent uncertainty in a model by random values sampled. This process is repeated about 1000 times (iterations) for each weight. The use of simulation is conducted under two primary constraints, i.e. maintaining the ordinal consistency and normalized weights (Diaby *et al.*, 2016)

- **SMARTER (SMART Exploiting Ranks)**

This method is developed to improve SMART weight which sometimes can be difficult because the respondent might not be confident to assign points to denote the relative importance of a criterion to the least important criterion. This method requires the respondent to order the criteria according to their importance. Afterwards, the SMARTER method allocates weights according to *rank-based weighting methods* (i.e. ROC, RS, or RR)(Riabacke *et al.*, 2012). Thus, the respondent does not have to assign points.

**Example**

(1) The respondent is asked to order the criteria based on their importance.

(2) The analyst derives the weights of criteria by using a weight processing method (i.e. RS, RR, ROC) in the previous section. Table 17 shows the weights from ROC method.

*Table 17.The example of SMARTER method with Rank Order Centroid*

| | Criterion 1 | Criterion 2 | Criterion 3 | Criterion 4 |
|---|---|---|---|---|
| **Rank position** | 4 | 3 | 1 | 2 |
| **Weights** | 0.0625 | 0.1458 | 0.5208 | 0.2708 |

(3) The total value of alternative $i$ $(MAV_i)$

$$MAV_i = (0.063 * Criterion\ 1) + (0.146 * Criterion\ 2) + (0.521 * Criterion\ 3) + (0.271 * Criterion\ 4)$$

Pairwise comparison

This approach involves the comparison of each criterion against every other criterion in pairs. It can be effective because it forces the DM to give thorough consideration to all elements of a decision problem (Marsh *et al.*, 2016).

- **Analytical Hierarchy Process (AHP)**

AHP utilizes pairwise comparison matrices of attributes. In this method, a decision problem is structured into a hierarchy of interrelated elements (Nikou *et al.*, 2015). Hence, AHP begins with the development of a conceptual representation of the decision in a hierarchy model. This hierarchy model comprises of the goal, the alternatives, the main criteria as well as sub criteria (depending on the problem) in order to evaluate how well the alternatives satisfy the goal in each level (Dolan, 2005). According to Nikou *et al.* (2015), the weight is estimated through pairwise comparisons of the importance of elements (i.e. criterion) within a level of hierarchy with respect to a higher level of the hierarchy. On the other words, pairwise comparisons are carried out among the criteria to define their relative importance in satisfying the goal, also among the alternatives to determine their relative abilities to fulfil the criteria (Dolan, 2005). The strength of their preference for a particular criterion over another criterion is denoted by 9-point semantic scale (Table 18) (Belton and Stewart, 2002). The outcomes are then compiled to generate a quantitative measure on a ratio scale that specifies how well each of the alternatives is expected to meet the goal (Dolan, 2005). Figure 13 shows a possible hierarchy for SHARP diet. The first level is the goal, the second level comprises of the SHARP dimensions, and the third level consists of the indicator(s) for each dimension. Then, the utility of diet alternative $i$ are judged according to this hierarchy.

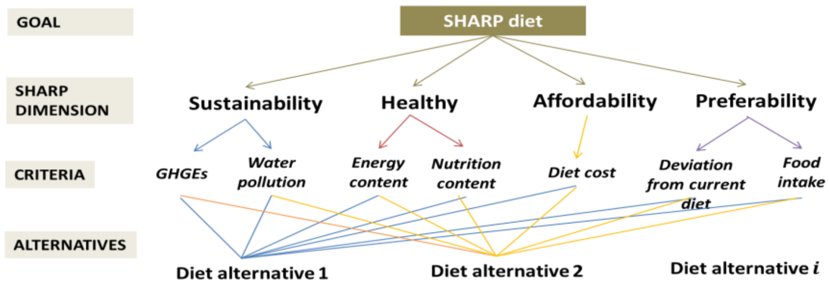


*Figure 13.SHARP diet's hierarchy*

The consistency ratio (CR) denotes the consistency of comparisons relative to a large number of purely random judgements. The value of '0' indicates entirely consistent judgements, while '1' implies random judgements. The acceptable consistency ratio is 0.1 or less (Belton and Stewart, 2002).

**Example:** (adapted from Belton and Stewart (2002))

(1) The analyst needs to determine the goal of the decision problem and the corresponding criteria to choose the best alternative in order to achieve the goal. The hierarchy can be enlarged by adding sub-criteria. However, in this example, we will not add sub-criteria for simplicity.

(2) The analyst also needs to determine some alternatives. In this example, there are three alternatives to be considered. The alternatives consist of all the four criteria. Given these elements of hierarchy, the visual representation is shown in Figure 14.

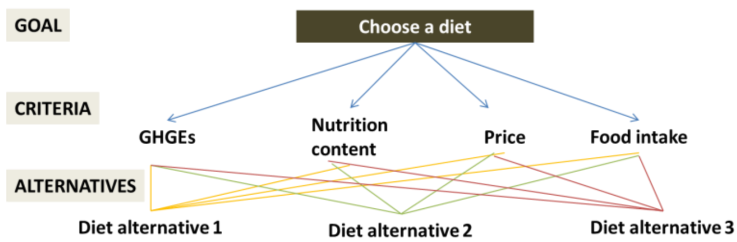


*Figure 14.The hierarchy model for the diet example*

(3) The respondent makes a pairwise comparison between the criteria. The respondent states how much more important the one criterion in comparison to the other criterion, using a 1-9 semantic

scale to express the strength of preference Table 18. The results will be in displayed in a matrix (Table 19, step 1). For example, **Criterion 3** (*Price*) is strongly preferred over **Criterion 1** (% of GHGEs reduction), thus receiving a value of 5.

*Table 18.The categorical scale for AHP (adapted from Belton and Stewart, 2001)*

| Scale | 1 | 3 | 5 | 7 | 9 | 2,4,6,8 |
|---|---|---|---|---|---|---|
| **Definition** | Equal importance | Moderate importance | Strong importance | Very importance | Extreme importance | Intermediate value |

(4) Continue with steps 1 and 2 with the remaining criteria (Riabacke *et al.*, 2012).

(5)The weights are normalized (step 2). The average weights are then calculated from the normalized weights. Table 19 illustrates the steps taken to calculate the final weights for each criterion.

*Table 19. The calculation step of AHP*

| | Step 1 | | | | Step 2 (normalization) | | | | Weights |
|---|---|---|---|---|---|---|---|---|---|
| **Criteria** | **1** | **2** | **3** | **4** | **1** | **2** | **3** | **4** | **w1+w2+w3+w4/4** |
| ***1. GHGEs reduction*** | 1 | 1/3 | 1/5 | 1/6 | 0.067 | 0.040 | 0.057 | 0.087 | 0.063 |
| ***2. Nutritional content*** | 3 | 1 | 1/3 | 1/4 | 0.200 | 0.120 | 0.093 | 0.130 | 0.136 |
| ***3. Price*** | 5 | 3 | 1 | 1/2 | 0.333 | 0.360 | 0.283 | 0.260 | 0.309 |
| ***4. Food intake*** | 6 | 4 | 2 | 1 | 0.400 | 0.480 | 0.567 | 0.521 | 0.492 |
| | 15 | 8.33 | 3.53 | 1.92 | | | | | |

The average weight across the rows is called *normalized Eigenvector*. This gives an importance order of = *Criterion 4 > Criterion 3 > Criterion 2 > Criterion 1.*

(6) The analyst computes principal eigenvalue ( $\lambda_{max}$ ) in order to calculate **Consistency Index** and **Consistency Ratio** (Equation 10).

$$\lambda v_p = \sum_{q=1}^{n} a_{pq} v_q \tag{10}$$

where, $a_{pq}$ = the entry in the $p$-th row and $q$-th column; $v_{p}$ = normalized eigenvector in $p$ −row; $v_q$ = normalized eigenvector in $q$ −column.

*To illustrate, the calculation starts by generating a new vector from the first row (GHGEs) = (1*0.063 + 1/3*0.136 + 1/5* 0.309 + 1/6*0.521) =0.257. This calculation continues until the last row. The $\lambda$ (eigen value) for each criterion is obtained by dividing those new vectors with their corresponding normalized eigenvector. For example, 0.257/ 0.063 = 4.07. Again, the analyst conducts this step until all eigenvector elements are used. The mean of these values will be the estimate of $\lambda_{max}$. The value of $\lambda_{max} \geq n$, where $n$ is the number of criteria; otherwise there might be some errors in calculation (Belton and Stewart, 2002).*

(7) The analyst computes consistency index (CI) with Equation 11.

$$CI = ((\lambda_{max} - n) \; / \; (n - 1)) \tag{11}$$

(8) The analyst calculates the consistency ratio (CR). Table 20 presents comparative values (CV) which vary in different matrix size.

$$CR = {CI}/{CV} \tag{12}$$

where $CV$ = comparative value whose value depends on the size of matrix/ number of criteria. $CR$ should be 0.1 or less to be acceptably consistent (Belton and Stewart, 2002).

*Table 20. The CV for AHP (taken from Belton and Stewart (2002))*

| **Size of matrix** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
|---|---|---|---|---|---|---|---|
| **Comparative value** | 0.52 | 0.89 | 1.11 | 1.25 | 1.35 | 1.40 | 1.45 |

(8) If the weights are found to be consistent, then the overall weight of each alternative can be calculated. The next step, the analyst should focus on *how each diet alternative compares to the other in terms of the certain criterion in question*. For example, when we create the "price" comparison, the respondent only compares the alternatives solely on the price, and nothing else. This rule applies to all criteria. Therefore, the analyst must create new matrices.
(9) The analyst conducts similar calculation as before to determine the weights for each criterion within alternatives. The hypothetical result from three alternatives is presented in Table 21.

*Table 21. Weights of diet alternative 1,2, and 3 in different criteria*

| | GHGE | Nutrition | Price | Food intake |
|---|---|---|---|---|
| **Diet Alternative 1** | 0.150 | 0.300 | 0.200 | 0.700 |
| **Diet Alternative 2** | 0.350 | 0.500 | 0.300 | 0.200 |
| **Diet Alternative 3** | 0.500 | 0.200 | 0.500 | 0.100 |

Overall value (MAV) for each alternative ($i$) is calculated based on an additive weighted function (Equation 13)

$$MAV_{(i)} = \sum_{j=1}^{N} w_j x_{ij} \qquad \forall \ i \in (1,\ldots K) \tag{13}$$

where, $N$ = the number of criteria; $w_j$ = weights of criterion $j$; $x_{ij}$ = weights of criterion $j$ within alternative $i$

*Example calculations for diet alternatives are shown below.*
$MAV_1$ = (0.063*0.150) + (0.136*0.300) + (0.309*0.200) + (0.492*0.700) = 0.456
$MAV_2$ = (0.063*0.350) + (0.136*0.500) + (0.309*0.300) + (0.492*0.200) = 0.282
$MAV_3$ = (0.063*0.500) + (0.136*0.200) + (0.309*0.500) + (0.492*0.100) = 0.262
The highest overall value is for alternative 1, thus diet alternative 1 is the best alternative.

- **Measuring Attractiveness by a Categorical-Based Evaluation Technique (MACBETH)**

The AHP method has been criticised due to the fact that the verbal expression in words might imply different meaning for different people thus it might lead to inconsistency when the semantic scale is converted to numeric scale (Riabacke *et al.*, 2012). MACBETH copes with the problem by incorporating both numeric and verbal statement simultaneously and continuous check by respondents.

**Example (**adapted from Bana E Costa *et al.* (2011) and Oliveira *et al.* (2012))
(1) The analyst has to define the criteria for each SHARP dimension also qualitative or quantitative descriptors for each criterion in order to create a performance scale. The descriptors must contain two limits: (i) **good** (*undoubtedly satisfying*); (ii) **neutral** (*neither satisfying nor satisfying*). The **good** level corresponds to a value of 100, while the **neutral** level to a value of 0. The number of descriptor levels might vary depending on the requirement of the criterion ((E Costa *et al.*, 2012, Oliveira *et al.*, 2012)). The level between good and neutral will have a value between 0 - 100. Table 22 gives an example of descriptors for *GHGEs* criterion.

*Table 22.The descriptor for GHGEs criterion in MACBETH*

| Levels | Descriptor : % of GHGEs reduction |
|---|---|
| **L1 (G=Good)** | 40% GHGEs reduction |
| **L2** | 30% GHGEs reduction |
| **L3** | 20% GHGEs reduction |
| **L4** | 10% GHGEs reduction |
| **L5 (N=Neutral)** | 0% GHGEs reduction |

(2) In order to generate a value function for each criterion, respondent is inquired to express MACBETH pairwise comparison judgement between performance levels of the respective criteria (Bana E Costa *et al.*, 2011, E Costa *et al.*, 2012). The respondent might use MACBETH 7-qualitative categories to indicate the difference in attractiveness (***very weak, weak, moderate, strong, very strong, extreme, or no***) (Oliveira *et al.*, 2012, E Costa *et al.*, 2012). Respondent might also give an answer to more than one category (i.e. *strong or very strong*, *moderate or strong*). Table 23 shows the example of a comparison matrix.

An example of question and answer: '*For L1 and L2, such that L1 is preferred L2, the difference in attractiveness between L1 and L2 is* ***strong***.'

*Table 23. The comparison matrix for descriptor levels*

| | L1 | L2 | L3 | L4 | L5 |
|---|---|---|---|---|---|
| L1 | | strong | strong | very strong | extreme |
| L2 | | | moderate | moderate-strong | very strong |
| L3 | | | | moderate | moderate-strong |
| L4 | | | | | weak-moderate |
| L5 | | | | | |

(3) The judgement process is assisted by software (i.e. M-MACBETH). The software is able to propose value scores by solving a linear programming problem and check the consistency of judgements. The interval between two levels varies depending on the judgement of attractiveness on step 2 as presented in Figure 15. Thus, the obtained value function might be piecewise linear (Oliveira *et al.*, 2012, E Costa *et al.*, 2012).

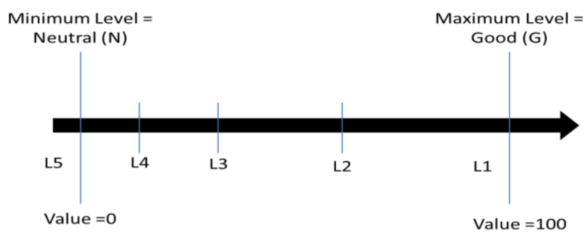


*Figure 15.An example of interval value in MACBETH*

(4) As a validation step of, the respondent might adjust the proposed values (Oliveira *et al.*, 2012). The value function can help to determine the value of criterion $j$ in each alternative $i$ ($v_{ij}$)

(5) The next step is to determine the weights for each criterion. Firstly, the respondent rank (in decreasing order of attractiveness) the importance of swings (from ***neutral level*** to ***good level***) of the four criteria. Table 24 presents the hypothetical result of the rank.

*Table 24.The hypothetical rank of criteria in MACBETH*

| *Indicator* | *Food intake* | *Price* | *% Fulfilment of Nutrition requirements* | *% of GHGE reduction* |
|---|---|---|---|---|
| *Rank* | 1 | 2 | 3 | 4 |

(6) The respondent qualitatively evaluates the attractiveness of each criterion' swing (Table 25). In the **neutral column**, the performance of the remaining three criteria is neutral. For instance, [GHGEs vs. NEUTRAL means *the swing* ***neutral to good*** *performance on GHGEs emission reduction compared to* ***neutral*** *performance on the remaining three criteria*]. The software will calculate the swings' value of each criterion ( $v_0([Criterion_j])$ ) using those judgements (e Costa et al., 2012).

An example of a question: '*How much more attractive is a swing from* ***neutral to the good*** *of sensory quality than a swing from* ***neutral to the good*** *of GHGEs reduction?'. The answer is* ***very strong***.

*Table 25. The comparison matrix for criteria's swing*

| | Food intake | GHGEs | Price | Nutrition | NEUTRAL |
|---|---|---|---|---|---|
| **Food intake** | | very strong | weak | moderate | very strong |
| **GHGEs** | | | moderate | very strong | weak |
| **Price** | | | | weak | strong |
| **Nutrition** | | | | | moderate |
| **NEUTRAL** | | | | | |

(5) Analyst can use the weights and the value of each criterion to calculate the overall attractiveness of each diet alternative (Bana E Costa *et al.*, 2011)

$$\forall\, j \in \{1,2, \ldots .N\} \text{ and } \sum_{j=1}^{N} v_o([Criterion_j]) = 100 \quad (14)$$

$$w_j = \frac{v_0([Criterion_j])}{100} \quad (15)$$

where, $w_j$ : the weight of each criterion; $v_0([Criterion_j])$ : the swing value of criterion $j$

(6) Respondents are always able to validate the weights proposed by the software. Moreover, sensitivity analysis of differences on participants' opinion regarding weights is also conducted until an agreement is reached and resulted in the final weight. In the end, the weight can be used in additive value model to obtain the overall value of each diet alternative (Oliveira *et al.*, 2012).

$$MAV(i) = \sum_{j=1}^{N} w_j . v_{ij} \quad (16)$$

where $MAV(i)$ : the total value of each diet alternative $i$; $v_{ij}$ : the value of criterion $j$ in alternative $i$

SWING Weighting

Generally, it aims to identify and assign 100 points to the criterion with the swing (range of performance) that matters most. This is followed by a pairwise comparison between this criterion and each of the others to determine the relative importance of swings in criteria, and correspondingly allocate the points between 0 and 100 (Marsh *et al.*, 2016).This acknowledges an element of randomness to observed choices due to the researchers' inability to identify all influences.

- ***Simple Multi-Attribute Rating Technique Swing (SMARTS)***

This method is a combination of SMART and swing. Instead of comparing the criteria itself, the respondents consider the swing of criterion from worst to the best level. Unlike SMART, the reference criterion is the most important criteria. The resulted weights from SMARTS are usually used to be further processed by other methods.

**Example** (adapted from (Riabacke *et al.*, 2012, Mustajoki *et al.*, 2005b).

(1) Respondent needs to consider all criteria (1-4) for food choice are at their worst consequence level. Respondent identifies the most important to change from worst to best level, assign 100 points to it. In our example, the most important criteria will be *price*, thus this criterion will get 100 points.

(2) Respondent continues with steps 1 and 2 with the remaining criteria (*food intake, nutritional content, and % of GHGEs reduction* respectively) and assigns fewer points that reflect the relative importance of the change compared to the change for the most important attribute. The hypothetical results are displayed in Table 26. The resulting weights are then normalized to sum to one (Riabacke *et al.*, 2012, Mustajoki *et al.*, 2005b).

*Table 26.The example of SMAR's calculation*

| CRITERIA | Respondents | | | | Calculation | |
|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **Mean weights** | **Normalized mean weights** |
| 1. *(% of GHGEs reduction)* | 50 | 20 | 40 | 40 | 37.50 | 0.142 |
| 2. *(Nutritional Content)* | 75 | 40 | 60 | 40 | 53.75 | 0.203 |
| 3. *(Price)* | 100 | 100 | 100 | 100 | 100.00 | 0.377 |
| 4. (*Food intake*) | 95 | 60 | 80 | 60 | 73.75 | 0.278 |
| | | | | Total | 100 | 1.000 |

(3) The value for alternative $i$ will be:

$$MAV_i = (0.142 * Criterion\ 1) + (0.203 * Criterion\ 2) + (0.377 * Criterion\ 3) + (0.278 * Criterion\ 4)$$

Scoring functions

This method assumes that value function is monotonically increasing or decreasing over the range of attribute measurement considered(Belton and Stewart, 2002).

- **Bisection**

With this method, the DM (respondent) is required to identify the point on the criteria scale (0-100) which is halfway, in value terms between two end points.

**Example**

For illustration, firstly we refer to one criterion, **% GHGEs reduction**

(1) Analyst defines the end points, for instance, the available alternative diets can reduce GHGEs emission from 0% to 50%. It is assumed that as the percentage of reduction increases from each diet, the value also increases.

(2) Analyst identifies the mid-point **value scale** (between 0-100 is then 50) that corresponds with the midpoint on the **%GHGE reduction scal**e (0-50%) according to the preference of respondent.

(3) Suppose that the mid-point of **%GHGE reduction** is at 10%, then the **mid-point of value scale** (score of 50) corresponds to 10% GHGEs reduction.

(4) The respondent should continue until the mid-point of **% GHGE reduction** for value scale of 25 and 75 are also discovered. The **%GHGE reduction** for those values should be within the range of 0-10% and 10-50%, respectively.

(5) Once all the mid-points are identified (i.e. 8% and 25% correspondingly), the sketch can illustrate the value function Figure 16.

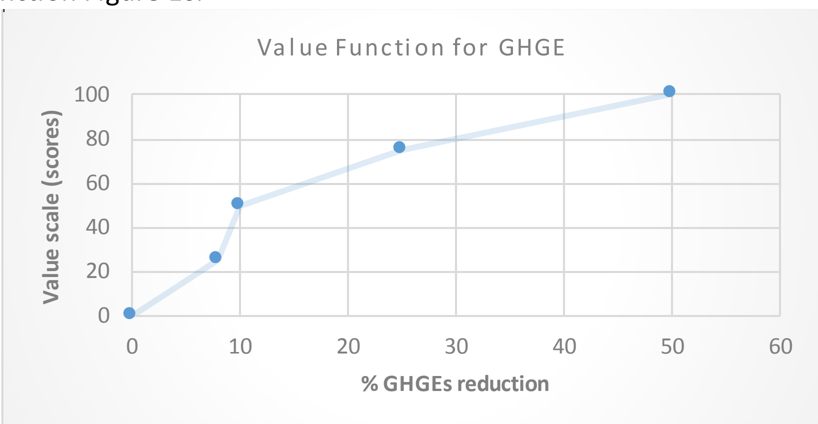


*Figure 16. Value Function for GHGEs*

(6) Since there are several criteria in diet preference model, then other value functions should also be made for other criteria. For example, value functions are also made for **nutritional content** (Figure 17) and **Price** (Figure 18) criteria. For price criterion, the value might increases when the price is lower because it will be more affordable for them.

(7) The final values can be calculated from the sum of all criteria in the alternative, depending on the levels of the criteria in the alternative. Assume that an alternative has the following criteria: Price (10 euros), Nutritional content (30%), GHGEs reduction (5%).
Based on the value function graphs, the total value is 50 + 35 + 10 = 95.

$$MAV(i) = \sum_{j=1}^{N} x_{ij} \;\; \forall \; i \in (1,\ldots K) \quad (17)$$

where, $x_{ij}$: the value of criterion on based on the level of criterion $j$ in alternative $i$

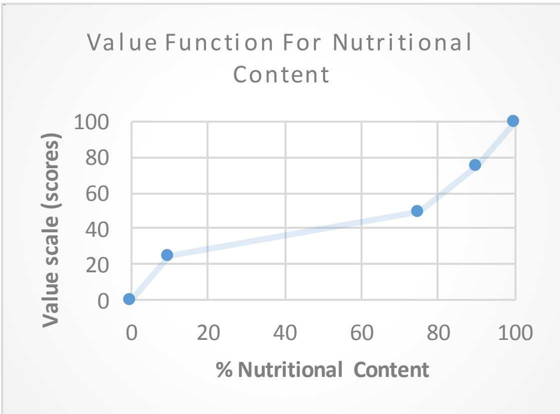


*Figure 17. Value function for Nutritional Content*

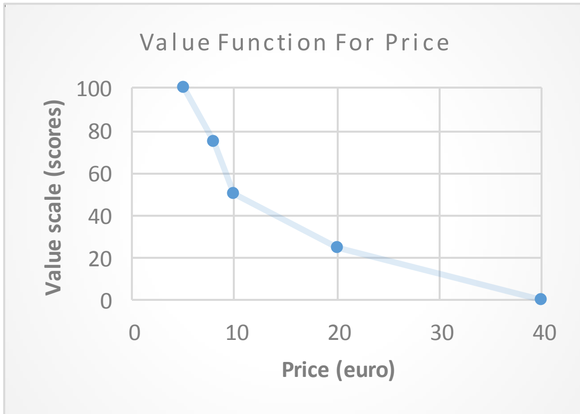


*Figure 18. Value function for Price*

## *Decompositional methods*

Similar to the previous section, an illustrative example is given to explain the mechanism of each method within the ***decompositional category***.

In this category, most of the methods belong to conjoint analysis and machine learning. Conjoint analysis (CA) is regularly used to predict consumer response regarding different product profiles. It works by determining the relative importance of these specified attributes/ criteria for consumers (Claret *et al.*, 2012, Asioli *et al.*, 2016). In general, there are two major CA categories: (i) Ranking-based conjoint analysis (so called ***acceptance-based approaches***); (ii) Choice-based conjoint analysis (so called ***preference-based approaches***) (Asioli *et al.*, 2016).

- Choice based

This choice-based refers to the approach where respondents are asked to make choices or rank from a series of sets of product profiles (alternatives). The methods presented here are part of Conjoint Analysis (CA). The goal of the conjoint analysis is to find the contribution of each criterion to the overall preference or utility (Hauser, 2007). This contribution is called as the 'part-worth' of the criterion (Hauser, 2007). Many articles also use DCE to denote CBCA. However, according to Louviere et al. (2010), they differ in the theoretical foundation. DCE is based on Random Utility Theory (RUT); while CA is not. RUT assumes that individuals are imperfect measurement tools. Therefore random components are incorporated to represent variability and differences in choices related to individuals. Due to this random component, utilities (or "preferences") are naturally stochastic. The analyst can forecast the probability that individual $n$ will opt for an alternative $i$, but not the exact alternative that individual will select. However, in the most identified articles about CBCA, they adapted the RUT (Cleemput *et al.*, 2014, Asioli *et al.*, 2016). Marsh *et al.* (2016) also refers CBCA and DCE as one method. Furthermore, a discussion in *sawtooth software* forum also stated that the difference is mainly in semantics (Sawtooth software, 2017). Due to those reasons, DCE and CBCA were considered similar in this study.

- **Choice-based conjoint analysis (CBCA) or Discrete Choice Experiment (DCE)**

This approach is widely used in academic and industry research (Asioli *et al.*, 2016). Principally, conjoint analysis method computes the consumers' trade-offs among multi-attribute goods or services by estimating the structure of consumer evaluation on a set of alternatives (called as *choice set*) comprising of predetermined combinations of attributes/ criteria' levels (referred to as *product profile or alternative*) (Asioli *et al.*, 2016). This choice-based approach (so-called ***preference-based***) is applicable when respondents are asked to make choices or rank from a series of sets of product profiles (alternatives) with varying level of criteria (Karniouchina *et al.*, 2009). This method is based on RUT (Baltussen *et al.*, 2007, Defechereux *et al.*, 2012). RUT estimates that individual always tries to maximize his/her utility when choosing among alternatives (Enneking *et al.*, 2007). For instance*, alternative 1* is chosen over *alternative 2* because $U_1 > U_2$.

**Example** (adapted from (Claret *et al.*, 2012, Asioli *et al.*, 2016, Defechereux *et al.*, 2012))

1) Firstly, the analyst decides upon the relevant criteria (attributes) and their levels. A focus group can be an option for those objectives (Claret *et al.*, 2012). Table 27 presents two possible levels for each criterion. The levels of the criteria depend on the target of our investigation. Furthermore, they need to be aligned with the consumers' acceptance.

*Table 27. Example of criteria and the corresponding levels for CBCA/DCE*

| No | Criteria | Levels | Description |
|---|---|---|---|
| **1** | **GHGEs reduction** | 5 % ; 10 % | The values represent feasible levels of GHGEs reduction. The higher reduction might |
| **2** | **Nutritional content** | inadequate; adequate | ***Adequate*** implies that the nutrition content is within the acceptable range of nutrition, while ***inadequate*** not. |
| **3** | **Price** | 20 euros; 30 euros | The price represents an average spending of meals per day. |
| **4** | **Food intake** | acceptable/ unacceptable | This criterion might focus on advisable amount vegetable and fruits. ***Acceptable*** means the amount ≥ 400 gram/ day (WHO, 2013), ***unacceptable*** if the amount < 400 gram/day. |

(2) The analyst uses an experimental design to combine those criteria and criteria levels into several sets of diet alternatives. *Full factorial designs* (includes all combinations of criteria and levels) in most cases are not appropriate due to a great number of feasible alternatives (Claret *et al.*, 2012). Thus, *fractional factorial design* (includes some parts of possible alternatives) is mostly preferred. The type of experimental designs is outside the scope of the study, but it can be found in other literature (i.e. *orthogonal design* by using a software *%Choiceff* macro (Asioli *et al.*, 2016) or SPSS (Claret *et al.*, 2012)).

Each set of diet alternatives is called choice task. The number of alternatives presented in each choice set depends on design. Usually, one choice task contains 2 or 3 alternatives. Figure 19 illustrates an example of a choice task comprising of three hypothetical diet alternatives with a 'no-buy' option.

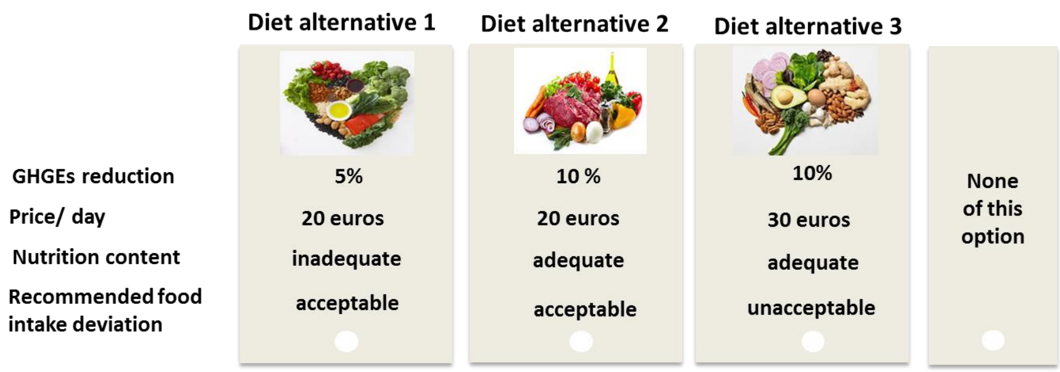


*Figure 19.An example of CBCA/DCE choice task*

(3) The alternatives are normally presented to the respondents with an explanation of their criteria and criteria levels as depicted in Figure 19. The respondent has to indicate their choice by choosing one of the diet alternatives for every choice set. If a 'no-buy' option is available, the respondent might also opt for that.

(4) The analyst performs a statistical analysis to generate estimated preference weights or choice-model parameters which are consistent with the obtained choices' pattern of respondents. For further information regarding the statistical analysis, some literatures can be looked at, such as discrete choice models (DCMs) (Enneking *et al.*, 2007), Mixed Logit (ML) (Asioli *et al.*, 2016), multiple regression analysis (Claret *et al.*, 2012), probit model (Karniouchina *et al.*, 2009), and multi-nominal logit (Næs *et al.*, 2011), conditional logistic regression model (Baltussen *et al.*, 2007).

(7) The analyst can use software to run the module and analyse the data. In principal, the model aims to estimate the coefficients that denote the preference parameters or marginal utility (so-called ***part-worth utilities of the criteria***) (Louviere *et al.*, 2010). Commonly, the total utility of each alternative is derived from *additive part-worth* (Defechereux *et al.*, 2012). The simple form of it is shown in Equation 18.

$$U_j = \sum_{i=1}^{I} \sum_{m=1}^{M} \alpha_{im} X_{im} \tag{18}$$

where , $U_j$ = utility of alternative $j$; $\alpha_{im}$ = part-worth utility for $m$-th level of criterion $i$, $X_{im}$ = dummy variable for $m$-th level of criterion $i$ (i.e. 0 or 1)

Table 28 provides a hypothetical result of part-worth utilities. According to Baltussen *et al.* (2007), each dummy variable ($X_{im}$) can be set equal to "1" when the qualitative level is present and set to "0" if it is not. The coding used in other literatures might differ.

*Table 28. The example of coded criteria levels and corresponding part-worth*

| | **GHGEs reduction** | | **Price/day** | | **Nutrition content** | | **Food intake deviation** | |
|---|---|---|---|---|---|---|---|---|
| **Level** | 5% | 10% | 20 euros | 30 euros | adequate | inadequate | acceptable | unacceptable |
| **Utility** | 0.30 | 0.35 | 0.50 | 0.30 | 0.20 | -0.80 | 0.78 | -0.70 |

Based on the data above, ***diet alternative 1*** will have a utility value of (0.30+0.50-0.80+0.78=0.78). Table 29 displays the calculation of the explainable component of diet alternative 1.

*Table 29. Example of calculation of $V_j$*

| | GHGEs reduction | Price/day | Nutrition content | Food intake deviation |
|---|---|---|---|---|
| Alternative 1 | 5% | 20 euros | inadequate | acceptable |
| CI value | *0.30* | *0.50* | *-0.80* | *0.78* |

- **Best-worst scaling (BWS)**

This method is an adaptation of 'pick-one option' CBCA/DCE method. Instead of only asking to choose the most preferred alternative, this method seeks additional information by asking respondents about the best and the worst elements of alternatives (Swancutt *et al.*, 2008, Al-Janabi *et al.*, 2011).The elements of the alternatives can be attributes (criteria), criteria level, and alternatives. There are three types of BWS according to those elements:(i) object case (criteria),(ii) profile case (criteria level),(iii) multi-profile (alternatives) (Mühlbacher *et al.*, 2016). The latter can extract more information than CBCA/DCE (Mühlbacher *et al.*, 2016). However, profile case is the most common approach in health-care (Al-Janabi *et al.*, 2011) and is less cognitively burdensome than CBCA/DCE because only one alternative is presented in every choice task (Swancutt *et al.*, 2008, Cheung *et al.*, 2016).

**Example**

The example will be given for profile-case. Initially, the analyst has to define the criteria (attributes) and their corresponding level for the experiment (similar to Table 27 in CBCA/DCE).

(1) The analyst creates the alternatives and choice sets using an experimental design. There are several methods available which can be studied from other literature (i.e. orthogonal array (Swancutt *et al.*, 2008), orthogonal main-effect design plans (Al-Janabi *et al.*, 2011)). An example of the choice task is in Figure 20.

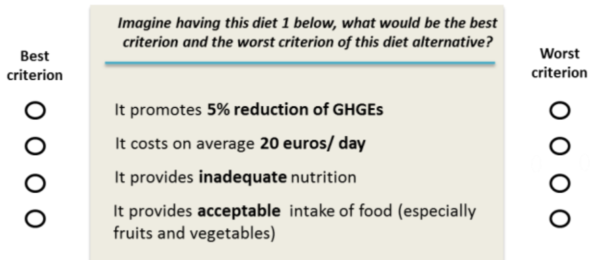


*Figure 20. An example of choice task in BWS-profile case*

(2) The respondents choose the best and the worst criteria within the alternatives for all the possible set of alternatives.

(3) Since BWS is an extension of CBCA/DCE, then the utility can be calculated with additive part-worth (Equation 20) in CBCA/DCE. Furthermore, the utility function can also be designed for each possible most/least important pair to measure the difference in utility between each pair chosen as best and worst. For example, the utility of selecting '5%' GHGEs reduction' as the best criterion and 'price of 20e' as the worst criterion can be defined as in Equation 19.

$$U_{(GHGEs_{5\%},price_{20e})j} = \left[\alpha_{GHGEs_{5\%}} * \left(1_{,if\ GHGEs\ is\ 5\%}\ ; 0_{otherwise}\right)\right] - \left[\alpha_{price_{20e}} * \left(1_{,if\ price_{20e}}\ ; 0_{otherwise}\right)\right] \quad (19)$$

| |
|---|
| where , $U_{(GHGEs_{5\%},price_{20e})i}$ = utility of selecting 5% GHGE reduction as the best criterion and 20e as the worst criterion; $\alpha_{GHGEs_{5\%}}$=part-worth utility of 5% GHGEs reduction; $\alpha_{price_{20e}}$= part-worth utility of diet price 20 *euros*. |

(4) In order to avoid over-specification, it is common to omit the least valued criterion level and consider it to be a base (Næs *et al.*, 2011). All utility estimates of the model can then be interpreted relative to the omitted level. Thus, the utility represents the additional utility of each criterion level over the base (omitted criteria) (Al-Janabi *et al.*, 2011). In BWS, it is usual that only one criterion level that needs to be omitted. While in DCE, it is required to fix one level of each criterion to simplify the model. It implies that all BWS' part-worth utilities are compared across all criteria levels, so the utilities in BWS have a common scale.

(5) In addition to that, ***count analysis*** can deliver intuitive pictures of individuals' choice by giving the best and worst frequencies (Al-Janabi *et al.*, 2011). The best-worst score can be built based

on the difference of *Total(Best) – Total(Worst)* (Mühlbacher *et al.*, 2016). For example, if the criterion level is chosen 10 times as best and 3 times as worst, then the score is 7. From the results, the ranking based on the BWS frequencies can be generated.

- **Potentially All Pairwise Rankings of all possible Alternatives (PAPRIKA)**

This method is a type of choice-based conjoint analysis (Golan and Hansen, 2012). This method proceeds by asking respondents to pairwise rank a series of possible alternatives, presented as a pair (dyads) in random order. The pairs of hypothetical patients are defined on two criteria at-a-time so that the respondents are forced to make a trade-off between the criteria (Golan and Hansen, 2012). This method minimizes the burden of respondents because when they answer a question, the method (via the software) eliminates all other possible questions that are implicitly answered as corollaries of those already answered (Golan and Hansen, 2012). It computes this by applying the logical property of "transitivity"; for example, if the respondent ranks hypothetical diet alternative "A" ahead of diet alternative "B" and also "B" ahead of diet alternative "C", then logically, "A" must be is ranked ahead of "C". Thus, a question pertaining to this third pairwise ranking would not be inquired (Golan and Hansen, 2012).

**Example** (adapted from Golan and Hansen (2012))
(1) The analyst determines the criteria and their corresponding levels (similar to Table 27)
(2) The analyst defines the dominated and undominated pairs. An 'undominated pair' is a *pair of alternatives where one is characterized by a higher ranked category for at least one criterion and a lower category for at least one other criterion than the other alternative* (Hansen and Ombler, 2008). The dominated pair alternatives are fundamentally pairwise ranked because one having a higher category for at least one criterion, but none lower for the other criteria ((Hansen and Ombler, 2008).
(3) The analyst creates some choice tasks containing two diet alternatives which initially differ in two criteria levels until all 'undominated pairs' are evaluated (Hansen and Ombler, 2008). An example of a choice task is in Figure 21.

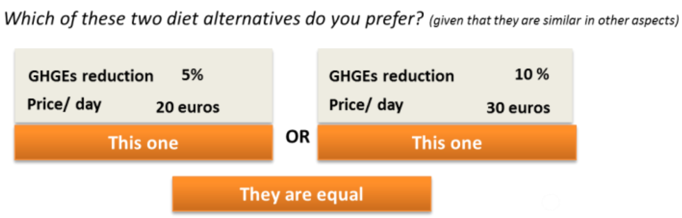


*Figure 21. An example of choice task for PAPRIKA*

(4) Respondents choose one option out of those the four options of answer (alternative 1, alternative 2, or equal)
(5) The software then uses linear programming (explained in detail in Hansen and Ombler (2008)) to calculate the '*point values*' based on the inequalities (>,<) and equalities (=) from the respondents' judgements. These point values reflect the relative importance of the criteria. The total point values of all criteria and criteria levels are 100.
(6) The analyst computes the value (total score) of each alternative from the summation of the point values as shown in Equation 20 (Golan and Hansen, 2012)

$$MAV_{(i)} = \sum_{j=1}^{N} x_{ij} \qquad \forall\ i \in (1,\ldots,K) \tag{20}$$

where, $x_{ij}$ = point values of criterion $j$ in alternative $i$

- Rating-based

2.1 Rating-based conjoint analysis

This approach is one of the main categories in conjoint analysis and also known as "**acceptance-based approaches**" (Asioli *et al.*, 2016). With this approach, respondents rate each diet alternatives based on their preference of different profiles (Karniouchina *et al.*, 2009, Asioli *et al.*, 2016).

**Example (**adapted from Asioli *et al.* (2016)**)**

(1) The analyst defines the criteria and their corresponding level to describe the diet alternative.

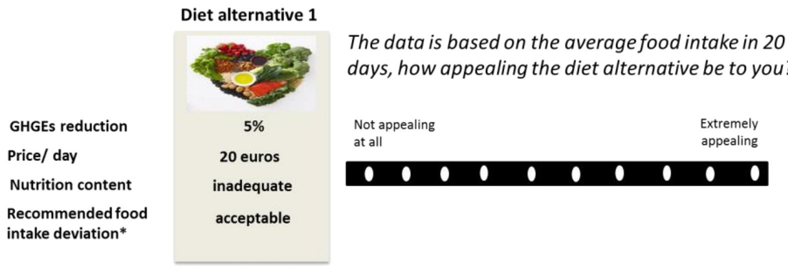


*Figure 22. An example of choice-task for RBCA*

(2) Similar to CBCA/DCE, the analyst creates the product profiles from experiment designs by using software. The aim is to limit the number of alternative profile presentations (Asioli *et al.*, 2016).

(3) The profiles of the alternatives are presented to the respondents with the description level of each criterion. The respondent provides an answer by rating them. An example of a choice task is in Figure 22. In this example, the answers are based on a 10-point scale, from 1 (not appealing at all) to 10 (extremely appealing).

(4) There is no fixed method to analyse the data. There same possible models, such as Mixed Model ANOVA (Næs *et al.*, 2011, Asioli *et al.*, 2016) or Ordinary Least Square (OLS) regression (Karniouchina *et al.*, 2009). We will not discuss these models in this research. In general, the model aims to estimate the part-worth utilities (Louviere *et al.*, 2010). The total utility can be computed as in CBCA/DCE (Equation 18).

- Learning based

- **Collaborative filtering**

This method is one of the approaches from Recommender systems (RSs). Recommender systems' work by suggesting items to users that are perceived to be desirable based on the analysis of their preferences (Elahi *et al.*, 2016). Collaborative filtering (CF) is one of the most successful approaches within RS (Elahi *et al.*, 2016). The essential assumption of CF is that if user *X* and *Y* rate *n* items similarly, or have similar behaviors (i.e. purchases), and hence will rate or act on other items similarly (Elahi *et al.*, 2016). In this case, users are considered to be similar if they co-rate similarly the items. This method gives recommendation an item/diet alternative that has not been rated yet by the target user, based on the highly rated items of similar users (Elahi *et al.*, 2016).

This method depends on numerous types of input: *implicit feedback* (i.e. purchase history) and *explicit feedback* (i.e. users' direct rating) (Koren and Bell, 2011). For *implicit feedback,* the greater the purchase intensity of a certain item indicates the higher the rating is. There are two well-known rating prediction algorithms; they are *neighbor-based approach* and *latent factor model* (Elahi *et al.*, 2016).

In a typical CF scenario, there is a list of *m* users and a list of *n* items (Koren and Bell, 2011). *Neighbor-based* approach concentrates on relationships between items or users (Koren and Bell, 2011). User-based approach generates an item's rating prediction for the target user according to how the similar users rate that particular item (Elahi *et al.*, 2016). *Latent factor models*, such as matrix factorization, include an alternative approach by transforming both items and users to the same latent factor space. The latent space explains ratings by characterizing both products and users on factors which are automatically inferred from user feedback (Koren and Bell, 2011).This method is notably employed in Amazon, Barnes and Noble, and Netflix because they are highly effective, easy-to-implement (Elahi *et al.*, 2016, Koren and Bell, 2011). Generally, latent factor models offer more accurate than neighbor-based models due to its ability to include various aspects of the data. Nevertheless, the simplicity of neighbor-based model makes this model more prevalent in the most literature (Koren and Bell, 2011).

**Example** (adapted from Elahi *et al.* (2016))

**A. Neighbour-based approach (user-based)**

(1) The analyst collects the data of users and their corresponding diet alternatives' ratings. The ratings might be deduced from the interactions among the users and the items within the system. In the application of *Amazon* or *Last.fm*, the users browse or buy certain items and sometimes provide ratings (Elahi *et al.*, 2016). In this case, diet alternative is considered as *item*. The hypothetical data is displayed in Table 30.

*Table 30. The user-item table*

| | **Diet alternatives** | | | |
|---|---|---|---|---|
| | $(I_1)$ | $(I_2)$ | $(I_3)$ | $(I_4)$ |
| **User 1** $(v_1)$ | 4 | ? | 5 | 5 |
| **User 2** $(v_2)$ | 4 | 2 | 1 | |
| **User 3** $(v_3)$ | 3 | - | 2 | 4 |
| **User 4** $(v_4)$ | 4 | 4 | - | - |
| **User 5** $(U_5)$ | 2 | 1 | 3 | 5 |

(2) The analyst can calculate the similarity or weight, which reflects distance, correlation, or weight, between two users, $u$ (active user) and $v$ (other users) or two items, $i$ and $j$. For this example, the similarity is calculated between users ($u$ and $v$). It is measured by computing the *Pearson correlation* (Equation 21) or other correlation-based similarities. Pearson correlation measures the extent to which two variables linearly relate with each other. For the user-based algorithm, the Pearson correlation between users $u$ and $v$ :

$$w_{u,v} = \frac{\sum_{i\in I}(r_{u,i}\, r_{v,i}) - \frac{\sum_{i\in I} r_{u,i} \sum_{i\in I} r_{v,i}}{n}}{\sqrt{\sum_{i\in I} r_{u,i}^2 - \frac{\left(\sum_{i\in I} r_{u,i}\right)^2}{n}}\sqrt{\sum_{i\in I} r_{v,i}^2 - \frac{\left(\sum_{i\in I} r_{u,i}\right)^2}{n}}} \tag{21}$$

where, the $i \in I$ summations are over the items that both the users $u$ and $v$ have rated; $w_{u,v}$= weight between active user $u$ and user $v$ ; $r_{u,i}$ /$r_{v,i}$= the rating given by active user $(u)$ and other users $(v)$ on item $i$; $n$ = number of items that both users active user $u$ and other users $v$ have rated

The example is carried out for user 1 $(u)$ and 5 $(v)$. Table 31 provides the data for Pearson calculation.

*Table 31. The data for Pearson calculation and example of calculation*

| | $r_{1,i}$ | $r_{5,i}$ | $r_{1,i}r_{5,i}$ | $r(1,i)^2$ | $r(5,i)^2$ |
|---|---|---|---|---|---|
| $(I_1)$ | 4 | 2 | 8 | 16 | 4 |
| $(I_3)$ | 5 | 3 | 15 | 25 | 9 |
| $(I_4)$ | 5 | 5 | 25 | 25 | 25 |
| **Total** | 14 | 10 | 48 | 66 | 38 |

$$w_{1,5} = \frac{48-(14*\frac{10}{3})}{\sqrt{(66-\left(\frac{14^2}{3}\right))*(38-\left(\frac{10^2}{3}\right)}} = 0.75$$

The values on the table can be computed from the equation to calculate the correlation between active user 1 and user 5. This computation continues for other users against the active user 1. The other results are 1 and 0 for $\boldsymbol{w_{1,2}}$ and $\boldsymbol{w_{1,4}}$ respectively.

(3) In the neighborhood-based CG algorithm, a subset of nearest neighbors of the active user is chosen based on their similarity with him or her. For this reason, a weighted aggregate of their ratings is used to generate predictions for the active user. In this example, we use a *weighted sum of others' ratings*. This technique computes a prediction of an active user $u$, on a certain item $(i)$ by taking a weighted average of all the ratings on that item according to Equation 22.

$$P_{u,i} = \bar{r}_u + \frac{\sum_{v \in V}(r_{v,i} - \bar{r}_v).w_{u,v}}{\sum_{v \in V}|w_{u,v}|} \tag{22}$$

Where: $\bar{r}_v$ and $\bar{r}_u$ = the average ratings for the user $v$ and user $u$ on all other rated items (excluding the item 2 $(I_2)$; $w_{u,v}$ = Pearson correlation between users (*u* and *v*). The summations are over all the users $v \in V$ who have rated item $i$.

The aim is to predict rating for active user $(u)$ 1 on item 2 $(I_2)$. Therefore, Table 32 displays the rating of active user and other users on item 2, as well as the average ratings as required by Equation 22.

*Table 32. The average ratings of users*

| Users ($u$ & $v$) | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $r_{v,2}$ & $r_{u,2}$ | ? | 2 | | 4 | 1 |
| $\bar{r}_v$ & $\bar{r}_u$ | (4+5+5)/3=4.67 | (4+1)/2 = 2.50 | (3+2+4)/3=3.00 | 4/1= 4.00 | (2+3+5)/3=3.33 |

$$P_{1,2} = \bar{r}_1 + \frac{\sum_v(r_{v,2} - \bar{r}_v).\, w_{1,v}}{\sum_{v \in V}|w_{1,v}|}$$

$$= \bar{r}_1 + \frac{(r_{2,2}-\bar{r}_2)w_{1,2} + (r_{4,,2}-\bar{r}_2)w_{1,4} + (r_{5,,2}-\bar{r}_5)w_{1,5}}{|w_{1,2}|+|w_{1,4}|+|w_{1,5}|}$$

$$= 4.67 + \frac{(2-2.5)(-1) + (4-4)(0) + (1-3.33)(0.756)}{1+0+0.756} = 3.95$$

Thus, the predicted rate for item 2 on user 1 $P_{1,2}$ is 3.95 $\approx 4.00$

**B. Latent factor models**

There is another approach called **latent factor models**. According to (Elahi *et al.*, 2016), one familiar example of this is matrix factorization. This method learns a vector of latent features (called *factors or criteria*) for each user and item (Elahi *et al.*, 2016). Every item factor represents how well an item or a user possess a certain latent aspect (Elahi *et al.*, 2016). For example, if the items are diet alternatives, the ***factors or criteria*** might be the indicators of SHARP dimensions or even another interpretable dimension. Moreover, user factor vectors measure the preference of the users for each factor. The objective of factorization algorithm is to divide original rating matrix R into two matrices, S and M, in a certain way so that their products can estimate the original matrix and forecast the missing ratings in the matrix.

$$R \approx SM^T \tag{23}$$

where, S = $|User| \times F$ matrix; M = $|Item| \times F$ matrix; $F$ = factors (which must be optimized)

The factor matrices (latent factor) are calculated by minimizing the sum of prediction errors and using regularization to avoid overfitting of the training data (Elahi *et al.*, 2016). There are some algorithms to conduct this, but it is not covered in this study. In principal, the algorithm can predict the missing ratings. The computed rating predictions are restricted to be between 1 to 5 in this case.

- **PageRank algorithm**

PageRank algorithm is a Web page ranking method that is used in Google search engine to determine which pages are most important. The basic idea is to define the importance of a Web page in terms of *importance* assigned to the pages hyperlinking to it. Initially, it is used in a Web page (Franceschet, 2011). However, this proposition can be exploited in different contexts (Franceschet, 2011). In the application of Webpage search, the web is perceived as a directed graph of pages connected by hyperlinks. Then, a random surfer begins from a random page and keeps clicking on successive links arbitrarily, going from page to page. The PageRank value of a page conveys the relative frequency the random surfer visits that page, assuming that the surfer goes on infinitely. The more time spent by the random surfer on a page, the higher the Page Rank importance of the page (Franceschet, 2011).

In the context of diet model, the directed graph can show which meal will likely be eaten after the consumption of a certain meal. Hence, the analyst can predict the diet pattern of the consumer by following the directed graph. In addition to that, the more frequent consumers consume a meal, the higher the importance (PageRank) of the meal. It could provide additional information during the diet optimization process, i.e. as a benchmark for the "sensibility" of a diet. An optimization model could generate a certain list of (optimal) food items, from which one could probably derive a set of meals. However, these meals might never appear in a sequence of a particular person's daily diet. This approach could, at least partially, correct that.

The PageRank of page $j$ is the sum of the PageRank scores of pages $i$ linking to $j$. Formally, the PageRank ($\pi_j$) of page $j$ can be described as follows (Franceschet, 2011)

$$\pi_j = \sum_i \pi_i h_{i,j} \tag{24}$$

| |
|---|
| where, $h_{i,j} = {}^{1}/_{q_i}$ (if there exists a link from page $i$ to page $j$, and $h_{i,j} = 0$ otherwise. It denotes as the probability that the random surfer moves from page $i$ to page $j$); $q_i$ = the number of distinct outgoing links of page; $\pi_i$= page rank of page $i$. |

Equation 24 depicts a basic algorithm of PageRank. An improved algorithm exists and is being used by Google. However, the algorithm is not in the scope of this study.

**Example**

(1) The analyst should obtain the data to infer the connections among meals. Figure 23 below depicts the directed graph of meal alternatives connected by links. So after consuming Meal W ($w$), respondents might also be interested in Meal Y ($y$), Meal Z ($z$), and Meal X (($x$) and so forth. The analyst summarizes this relationship of the ingoing and outgoing links from the directed graph into equations (26,27,28,and 29).The out-flows of each alternative depend on how many links it has. For Meal W ($w$), it has 3 out-links, then each links equal to $w/3$.

(2) Total probability of all connection is 1, so then

$$w + x + y + z = 1 \tag{25}$$

$$w = \frac{z}{3} \tag{26}$$

$$x = \frac{z}{3} + \frac{w}{3} \tag{27}$$

$$y = \frac{x}{1} + \frac{z}{3} + \frac{w}{3} \tag{28}$$

$$z = \frac{y}{1} + \frac{w}{3} \tag{29}$$

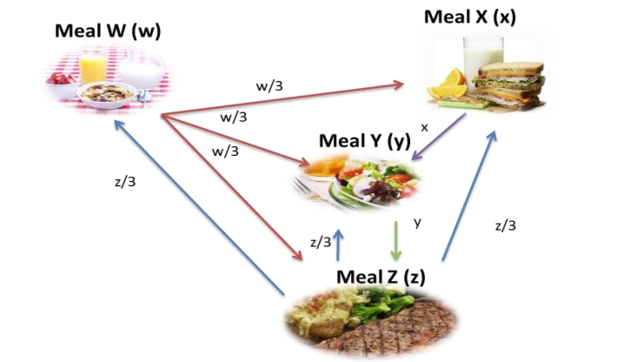


*Figure 23. The directed graphs of diet pattern*

(2) The analyst computes the variables in terms of a reference variable. In this case, $z$ is the reference variable.

$$x = \frac{z}{3} + \frac{z}{9} = \frac{4z}{9}$$

$$y = \frac{4z}{9} + \frac{z}{3} + \frac{z}{9} = \frac{8z}{9}$$

(3) The analyst processes the variables to find the importance of reference $z$

$$1 = \frac{z}{3} + \frac{4z}{9} + \frac{8z}{9} + z$$

$$\frac{24z}{9} = \frac{3z}{9} + \frac{4z}{9} + \frac{8z}{9} + \frac{9z}{9}$$

$$24z = 9$$

$$z = 3/8$$

(4) Analyst calculates the importance of other alternatives

$$w = \frac{1}{8} \qquad x = \frac{1}{6} \qquad y = \frac{1}{3}$$

(5) The analyst can use the values to estimate the diet pattern of consumer based on the probability of the links. For instance, after consuming Meal W, respondent has a similar probability of choosing among Meal X, Y, and Z. Suppose that the respondent chooses Meal Y, the afterwards respondent will have higher tendency to consume Meal Z. The importance of a certain meal can be calculated from the *additive weighted model* as mentioned in Equation 25.

### *4.5 The main components of the methods*

Table 33 and 34 provide the summary of the main components of each method. There are three main elements, namely *extraction, representation, and interpretation* as mentioned by Riabacke *et al.* (2012). The results show that rating and ranking are common extraction methods. However, it is observed that the extraction of respondents' input for machine learning is not merely limited to rating or ranking as the other methods. Another type of information input such as purchases history (i.e. in CF) can also be utilized (Koren and Bell, 2011). The collected information is then used to uncover the hidden pattern of data. The extraction of input is usually carried out through two ways, *joint procedure* (the criteria are compared all together) and *pairwise procedure* (the criteria are compared in pairs). Conjoint stimuli are common in the conjoint analysis where the alternatives are the combination of all criteria (BWS, RBCA, CBCA), and pairwise dyads are when only two criteria are shown in each alternative (PAPRIKA).
Strikingly, in terms of representation, only pairwise methods (AHP and MACBETH) which use semantic descriptive to represent DM's input. The rest of the methods utilize point estimates (cardinal), ordinal, or interval points. The most commonly used utility model among the methods is an *additive function*. The addictive function has an advantage of being more communicative and

clearer for analysts (Marsh et al., 2016). There are two common types of addictive function, namely weighted *additive* and *additive part-worth utility*. The weighted additive is denoted by multiplication of weights and values of criteria. The weighted additive is frequently observed in compositional method, while the additive part-worth utility is in decompositional method (i.e. BWS, CBCA, RBCA). Normalized weight is a typical interpretation procedure in MCDA when the DM's input is denoted as point estimates. When the DM's input is indicated as ordinal, the rank-based weighting methods (i.e. RS, RR, ROC) are necessary to convert the rank into *surrogate weights*.

To summarize, there are three major components of preference elicitation which differentiates one method to another. The common extraction methods are rating and ranking, while the typical utility model is an additive function. Point estimates (cardinal) and ordinal are frequently used to represent DM's input. The semantic descriptive representation is usual in pairwise comparison sub-category (i.e. AHP, MACBETH).

*Table 33.Main components of compositional methods*

| COMPOSITIONAL | | | | |
|---|---|---|---|---|
| **Methods** | Extraction | Representation | Interpretation | |
| | | | Utility model | Procedure |
| **Scales** | Rating, joint procedure | Point estimates | Weighted additive | Normalized weights |
| **Point allocation** | Rating, joint procedure | Point estimates | Weighted additive | Normalized weights |
| **AHP** | Ratio-scale, pairwise procedure | Semantic descriptive 9-point ratio scale | weighted additive | Eigen vector |
| **MACBETH** | Ratio-scale, pairwise procedure | Semantic descriptive 7-point ratio scale | Weighted additive | Linear programming |
| **SMARTS** | Ranking and Ratio-scale, pairwise comparison to the most important | Point estimates | Weighted additive | Normalized weights |
| **Interval SMART** | Ratio-weight procedure | Interval endpoints | Weighted additive | Linear programming |
| **SMART** | Ranking and Ratio-scale, pairwise comparison to the least important | Point estimates | Weighted additive | Normalized weights |
| **Ranking** | Ranking, joint procedure | Ordinal | Weighted additive | Rank sum, Rank order centroid, Rank Reciprocal |
| **Bisection** | Half point between certain increments | Interval midpoints | Additive value | Value function |
| **SMARTER** | Ranking, joint procedure | Ordinal | Weighted additive | Rank Order Centroid |

*Table 34.Main components of decompositional methods.*

| DECOMPOSITIONAL | | | | |
|---|---|---|---|---|
| **METHODS** | EXTRACTION | REPRESENTATION | INTERPRETATION | |
| | | | UTILITY MODEL | PROCEDURE |
| **PAPRIKA** | Ranking, pairwise dyads comparison | Ordinal | Additive part-worth utilities | Linear programming |
| **BWS** | Ranking (best & worst), conjoint stimuli | Ordinal | Additive part-worth utilities | Weighted Least Squares or Conditional Logistic Regression |
| **RBCA** | Rating, conjoint stimuli | Point estimate | Additive part-worth utility | Mixed Model ANOVA, Parameter Logit model (RPL), Latent Class Logit model (LC), Ordinary Least Squares (OLS) regression analysis |
| **CBCA** | Ranking, conjoint stimuli | Ordinal | Additive part-worth utility | Binary logistic Regression model, conditional logistic (CL) |
| **Collaborative filtering** | Rating or purchase /consumption history | Point estimate | Weighted additive | Pearson correlation; Matrix factorization |
| **PageRank** | Purchase /consumption history | Point estimate of probability | Weighted additive | PageRank algorithm (with probability) |

### 4.6 Advantages and disadvantages of reviewed methods

The advantages and disadvantages of the reviewed methods were based on the rational evaluation criteria of Table 7. The complete result is available in Appendix III. The data analysis for snowball methods Table 35 summarizes the advantages and disadvantages of the methods. The comprehensive discussion is presented as follows:

**a. Weights typology**

The weight typology depends on whether during the elicitation task, they generate scaling constants (*trade-off*) or merely indicate the importance of criteria (importance coefficients) (Marsh *et al.*, 2016, Cinelli *et al.*, 2014). In preference elicitation, if the method considers trade-off, it might increase the precision of the elicitation since it may represent the actual *trade-off* value from the decision maker (Marsh *et al.*, 2016). The value trade-offs are important in multiple criteria problem to indicate the relative desirability of achievement on each criterion in comparison to the others (Riabacke *et al.*, 2012). These qualifications are best satisfied by the swing weighting, decompositional approaches (Marsh *et al.*, 2016), AHP (Cinelli *et al.*, 2014), MACBETH, and point allocation. Methods which do not consider trade-off tend to give similar weights to the criteria (i.e. scales) (Marsh et al., 2012).

**b. Treatment for uncertainty**

This parameter relates to the capability of methods in handling uncertain, imprecise or missing information (Cinelli et al., 2014). The uncertainty in food system might arise from the price of foods. Therefore the existence of uncertainty treatment is advantageous. There are various treatments of uncertainty found in the methods. In the compositional group, there are various treatments to handle imprecise judgement. In the pairwise method (i.e. MACBETH), the respondents are able to indicate the difference of attractiveness with more than one category scale (i.e. *strong to very strong*). Interval SMARTS also provides interval value for defining the range of preference. The ranking method does not demand respondents to specify precise preference weights, yet it asks the order of preferences. For scoring function (i.e. bisection), the preference weights are not static but monotonically increasing or decreasing over the range of criterion measurement (Belton and Stewart, 2001). It is useful to capture the actual preferences because the level of satisfaction might vary according to the level of criteria. For example, the importance of GHGE reduction from 20 % to 40% might be more valuable compared to from 50% to 70% because the latter diet might be less acceptable. However, scales and point allocation do not accommodate imprecise judgement or information since they require a precise judgement from respondent.

In decompositional choice and rating based methods, the uncertainty of answers usually is helped by providing “no-buy” option (i.e. RBCA, CBCA, and BWS). Thus, the DM does not have to choose when he/she feels uncertain about their decision. One of the uncertainties that might occur is the variety of time (i.e. winter, fall, spring, summer) which might alter the composition of the diet. The CF methods (i.e. latent factor models) handles this uncertainty by taking into account the ‘variety of time’ in the algorithm (Koren and Bell, 2011)

In conjoint analysis methods (CBCA,RBCA, BWS), the missing value can also be handled yet it will require complicated procedures to keep the validity of analysis without excluding the respondent from analysis (Nikou *et al.*, 2015). However, PAPRIKA can tackle missing value by estimating from other pairwise comparisons. Machine learning methods might involve direct rating from users to minimize missing information (Elahi *et al.*, 2016) because too many missing values might decrease the accuracy of diet recommendation. Meanwhile, due to the simple design of compositional methods, these methods might cope with missing value more easily than decompositional. As an example from Nikou *et al.* (2015), AHP can tackle the missing values easily by relying on the consistency measure to estimate missing preference values in the pairwise comparison matrices (Nikou *et al.*, 2015). The method to handle this problem is usually incorporated in commercial software implementations of AHP. The similar case might be applicable for MACBETH as well since it includes continuous consistency check.

**c. Robustness**

The method is considered as robust when the addition or deletion of an alternative does not affect the classification or ranking of others. An example is the occurrence of *rank reversal* – reversal in the ranking (Cinelli *et al.*, 2014). Rank reversal mostly occurs in a pairwise comparison, (i.e. AHP) , not in a condition where all criteria are considered at once (i.e. decompositional methods, ranking, scales, swing weighting) (Cleemput *et al.*, 2014). Rank reversal is ascribed to the poor quality of the information available (Cinelli *et al.*, 2014) and inconsistency of pairwise-comparison (Felli *et al.*, 2008). It occurs because AHP does not need the preferences to be *transitive* (if *x* is preferred to y, and *y* is preferred to *z*, then *x* must be preferred to *z*) (Marsh et al., 2016). In AHP, the consistency check is conducted through the calculation of consistency ratio (CR). However, there is still a chance of consistency may occur because there is 10% of error tolerated. PAPRIKA method is also based on pairwise-comparison, but it always applies transitivity check. Therefore PAPRIKA is considered robust. MACBETH incorporates continuous consistency check to the acquired answers (E Costa *et al.*, 2012, Oliveira *et al.*, 2012). Thus, the quality of data is higher than AHP. Hence, it should be less prone to rank reversal than AHP. Nonetheless, the consistency check in MACBETH seeks for high involvement from DM. Thus, it might cause a drawback in terms of ease of use. SMART does not specify an upper limit for its judgement. Hence the obtained rating from the similar person might differ significantly if the method is applied twice, thus it might impair its consistency (Riabacke *et al.*, 2012). For machine learning techniques, as long as the data is sufficient and stable to generate reliable recommendation, then rank reversal is not a problem.

**d. Ease of use**

This parameter particularly deals with the cognitive burden implied to the decision maker as respondent. The number of required input should be tolerable to alleviate the burden for the decision maker(s) and the complexity of data analysis. Simple compositional methods, such as ranking and scales are perceived to be easy (Bottomley and Doyle, 2001). Point allocation is somewhat more cognitively demanding than those two other methods because the respondent has to keep track on the amount of points that they have distributed (Bottomley and Doyle, 2001). BWS (especially profile case) is potentially easier than DCE/ CBCA (Swancutt *et al.*, 2008) because the respondent does not have to consider a profile, not multiple profiles, for each choice task.

The conjoint stimuli are also considered to be easy, as long as the number of criteria is not more than 8 (Nikou *et al.*, 2015); otherwise, the number of profiles will increase and result in higher cognitive burden. These rules also apply to other compositional and decompositional approaches, except machine learning methods (i.e. CF, PageRank) because they can gain the data from transaction database (as implicit feedback). If the rating data is limited, user profile information (i.e. gender, age, education, etc.) can be employed as user's similarity information (Adomavicius and Tuzhilin, 2005). This method is called as demographic filtering.

The availability of tools to implement the method, manage the information, and show the results are highly useful in analysing the data and support the analyst (Cinelli *et al.*, 2014). Most of the methods, specifically from the decompositional group, are supported by software for analysing the data. Complicated compositional methods also have software support, such as AHP and MACBETH. There was not much information found regarding the software for basic compositional methods (i.e. ranking, rating). There was only one literature revealed that Excel could be used to a simple descriptive statistic in analysing the rating data (Goetghebeur *et al.*, 2012). The CF and Page Rank can automatically analyse the obtained information and subsequently derive the recommendation or pattern (Elahi et al., 2016; Franceschet, 2011).

**e. Learning dimension**

This learning dimension is related to possibility of re-evaluating results if new information becomes available (e.g. alternatives or criteria). Thus, it is not required to re-run the software and obtain independent results (Cinelli *et al.*, 2014). This is affected by the structure of the methods, whether it is single or multi-stage approach (Nikou *et al.*, 2015). The multi-stages approach is commonly attributed to the preparation stage, especially in choice-based and rating-based decompositional approaches. They demand the generation of design for alternatives and choice task. Thus, if a new criterion is added, the whole process has to be redone. For MACBETH, the analyst also need to define the descriptor first and generate value function before the pairwise comparison (E Costa *et al.*, 2012). Hence, this is also a multi-stages approach. One-stage or simple method (i.e. direct rating, ranking) will provide better learning dimension.

In machine learning methods, the addition of new user or alternatives will cause a *cold-start* problem. This problem takes place when the system has not yet acquired enough data to generate reliable recommendations or to capture the diet pattern (Elahi *et al.*, 2016).

**f. Sample size**

Sample size indicates the potential number of respondents needed also has to be considered (Nikou *et al.*, 2015). Smaller sample size might be easier in terms of practicality, but not necessarily with reliability. There is no clear rule about the amount of sample size needed for each method. In conjoint analysis (i.e. RBCA, CBCA/DCE), experience from the literature shows that above 100 respondents, a sample is good enough (Nikou *et al.*, 2015, Asioli *et al.*, 2016). In some cases (Van Wezemael *et al.*, 2014, Saito and Saito, 2013), the respondents even reached >500 to as representatives of the population. In BWS, Al-Janabi *et al.* (2011) mentioned that the most recently published study achieved significant results with 30 individuals per subgroup of interest. The variation of sample size was also observed with PAPRIKA. There was one study using 61 respondents (Golan and Hansen, 2012), but there was a study which involved only 8 experts (Johnson *et al.*, 2014).

According to Pöyhönen and Hämäläinen (2001) which studied the application of AHP, SMART, SMARTER, SWING, Direct Rating, even though there is no consensus about the sample size, the number of respondents less than 50 is considered to be small and might cause unreliable results. It will help to observe when the final results become stable. From their experiment, 150 respondents were required to achieve stable results (Pöyhönen and Hämäläinen, 2001). The information of sample size needed in Bisection (Belton and Stewart, 2002) and Interval SMARTS (Mustajoki *et al.*, 2005b) were not even given. Similarly, the literature also did not provide an exact guidance for minimal sample size for BWS application.

According to De Bekker-Grob *et al.* (2015), the sample size for CBCA/DCE can be calculated by considering several factors (i.e. model to be used, significance level, statistical power level, etc.). Cheung *et al.* (2016) mentioned that future research is needed to define sample size calculations based on the desired statistical power as what have been done in CBCA/DCE.

The main difference of sample size was observed in machine learning techniques. In practice, collaborative filtering and PageRank are used to evaluate a large product set which can be up to millions (Franceschet, 2011) in Netflix and Google search engine respectively. To illustrate a case in Netflix, on average a movie receives 5600 ratings, while a user rates 208 movies (Koren and Bell, 2011). Therefore, the sample size needed for these methods is enormous.

**g. Context of use**

The context of use of preference elicitation methods might be different. Since CA and its derivatives (i.e. BWS, DCE/CBCA, and RBCA) were developed mainly in the context of marketing, they are mostly engaged in understanding consumers' needs. Their primary objective is to conclude the most important factors to focus on when designing a new product or service (Nikou *et al.*, 2015). At this point, the familiarity of use is not applicable because there is no existing product in the market yet.

One illustration from the study of (Annunziata and Vecchio, 2013) who did research to explore consumers evaluation of four attributes of probiotics functional foods. The functional foods did not exist yet in the market; they just wanted to understand the required attributes for functional food. Thus, experience with that particular product was not necessary. Several other examples are to investigate: lamb meat attributes (Gracia and De-Magistris, 2013), functional food attributes (Annunziata and Vecchio, 2013), computer attributes (Karniouchina *et al.*, 2009), type of information in food label (Caputo *et al.*, 2013) or in nutrition label (Van Wezemael *et al.*, 2014) .
At the same time, originating from MCDA, the applications of compositional methods, such as AHP (Nikou *et al.*, 2015) and MACBETH (Mustajoki *et al.*, 2005b, Oliveira *et al.*, 2012), can be mainly described as situations when there is a number of alternatives available and by determining the importance of different characteristics, we intend to choose the best alternative in the context of product/ policy/ treatment development. On the other words, after the design phase, there are still several possible prototype products, and we aim to choose the best one. At this stage, preference elicitation process puts a higher importance on the experience/familiarity with a product.
An example is from Swancutt *et al.* (2008) who tried to identify patients' preferences for aspects of appointments within the colposcopy service and to make suggestions for service improvement by using BWS. For this case, familiarity or experience of respondents with the service is highly important to be able to provide relevant input towards the colposcopy service.
Based on the results of the review, the common *context-of-use* of PAPRIKA was also similar to AHP and MACBETH. The literature showed that it was used to determine which health technology should be funded (Golan and Hansen, 2012) or which patient should be prioritized (Hansen and Ombler, 2008).
Machine learning methods (i.e. CF and PageRank) also have a purpose to select the most suitable alternative based on the learning data and to recommend it to the user. For this purpose, machine learning methods demand input from other users who are familiar with the product to predict the most suitable product for another user (Elahi *et al.*, 2016).
This distinction does not necessarily mean that certain method is not suitable to select the best alternative or to determine general consumer needs. This rule only refers to the difference in the most common application areas (Nikou *et al.*, 2015).The bottom line is that the knowledge-level of the consumers should be aligned with the context of use. Furthermore, one should choose the most appropriate method to evaluate consumer preferences based on this context.

To sum up, there are several advantages and disadvantages derived from the results. In terms of sample size, machine learning technique might require more data than other methods to provide reliable recommendations. However, there is no exact rule to determine the sample size for almost all the methods which are representative for the population. Most of the methods are easy when the number of criteria and criteria level are limited, but become cognitively burdensome for DM when they are increased. This is particularly true for techniques under pairwise comparison, scoring function, conjoint analysis methods. AHP is highly prone to rank reversal because it does not require transitivity. MACBETH should be less prone to rank reversal.
If the analyst wants to add more criteria or criteria level into the model, multi-stages methods (i.e. conjoint analysis and pairwise comparison) are not handy for conducting re-evaluation. With regard to uncertainty treatment, the results indicated that most methods use a value interval, categorical scale, or ‘no-buy’/indifferent option to deal with imprecise judgement. Additionally, missing value can also be handled by the methods with various techniques. Lastly, in terms of *context-of-use*, it should be noted that the knowledge level of the DM should be aligned with the context of elicitation. MCDA and machine learning methods are commonly applied in evaluating or screening of product and service, while conjoint analysis methods are typically utilized to determine user’s needs.

*Table 35. The advantages and disadvantages of compositional and decompositional methods*

| ***Methods*** | **Weights** | **Treatment for** | **Robustness** | **Context** | **Ease of use** | **Learning** | **Sample** |
|---|---|---|---|---|---|---|---|

| | typology | uncertainty | | of use | | dimension | size |
|---|---|---|---|---|---|---|---|
| *Compositional* | | | | | | | |
| *Scales* | **no trade off** | **no treatment** | no rank reversal | Screening of product /service | easy | easy to re-evaluate | < 100 |
| *Ranking* | **no trade off** | imprecise valuation | no rank reversal | Screening of product /service | easy | easy to re-evaluate | >50 |
| *Point Allocation* | trade off | **no treatment** | no rank reversal | Screening product & understand users needs | easy ; (-) keep track on points | easy to re-evaluate | >50 |
| *AHP* | trade-off | -easy to handle missing value | **rank reversal** | screening of product or service | -easy for a few criteria -software support (i.e. AHP) | **multi-stages** | >50 |
| *Interval SMART* | trade-off | Imprecise judgement (interval) | no rank reversal | screening of product or service | -easy for only a few criteria -software support (i.e. WINPRE) | easy to re-evaluate | NA |
| *MACBETH* | trade-off | -imprecise judgement -easy to handle missing value | **less prone to rank reversal, (consistency check)** | screening of product or service | **-high effort for DM -software support (i.e. M-MACBETH)** | **multi-stages** | <100 |
| *Bisection* | **no trade-off** | Imprecise judgement (interval) | no rank among criteria | Screening of product /service | **high effort for DM** | easy to re-evaluate | NA |
| *SMARTS* | **no trade-off** | NA | no rank reversal | Determine users' needs | easy | easy to re-evaluate | >50 |
| *Decompositional* | | | | | | | |
| *PAPRIKA* | trade-off | -indifferent option -category scale -easy to handle missing value | transitivity check, prevent rank reversal | Screening of product or service | -minimize question based on corollaries -software supports (1000Minds) | **multi-stages** | < 100 |
| *BWS* | trade-off | -'no-buy' option -category scale -difficult to handle missing value | no rank reversal | Determining user needs | - less cognitive burden - software support (i.e. SPSS) | **multi-stages** | min. 30 |
| *RBCA* | trade-off | -'no-buy' option -category scale -difficult to handle missing value | no rank reversal | Determining user needs | -easy for a few criteria, 'no-buy' option -software support (i.e. SPSS) | **multi-stages** | >100 (rule of thumb) |
| *CBCA/DCE* | trade-off | -'no-buy' option -category scale -difficult to handle missing value | no rank reversal | Determining user needs | -easy for a few criteria, -'no-buy' option -software support (i.e. SAS) | **multi-stages** | >100 (rule of thumb) |
| *Collaborative filtering* | trade-off | -uncertainty factor in the algorithm -use direct rating to handle missing value | no rank reversal | Screening of product or service | -survey is not needed, data could be from history -computer support | **cold start** | **>> 100** |
| *PageRank* | trade-off | **NA** | no rank reversal | Screening of product or service | -data from the database -computer support | **cold start** | **>>100** |

*note: dark grey box for 'disadvantages,' white box for 'advantages,' light grey for 'NA.'*

### 4.7 Conditions of using the preference elicitation methods in SHARP diet

The purpose of this study is to review the most representative preference elicitation methods and provide a recommendation about methods that might be suitable for SHARP diet modelling. Basically, the reviewed preference elicitation methods could also be applicable for SHARP diet, depending on the conditions as follow:

**- The identification of indicators for SHARP diet**
According to the previous research (Faramitha, 2016), the most important step in the establishment of SHARP diet is the determination of performance SHARP indicators**.** The validity of the research findings depends on the analyst's ability to correctly specify the relevant diet criteria and levels. Each dimension of SHARP might be represented by multiple indicators or criteria. To simplify the MP diet, analyst usually only considers the most representative criteria for each dimension. *For a limited amount of criteria*, MCDA and conjoint analysis (i.e. compositional methods, choice-based, and rating-based decompositional methods) might be more appropriate (Mühlbacher *et al.*, 2016, Cheung *et al.*, 2016). In the case of RBCA, BWS, and DCE/CBCA, Nikou *et al.* (2015) specified that the preferable number of criteria is at most 8, or else the survey will be too cognitively demanding for respondents because the respondents have to examine more alternatives. To come up with criteria and criteria levels, a focus group must be carried out prior to the survey. Thus, the establishment of criteria and criteria levels might also require extra effort from respondents. The advantage is that analyst could derive the importance value (i.e. weights or part-worth utilities) of each criterion or criteria levels. The drawback of this limitation is that some other relevant criteria which might be influential on the weights, yet not included in the model. As indicated by , the exclusion of some criteria might alter the outcome of the model(Goetghebeur *et al.*, 2012). For example, the price of diet is an important criterion for DM, but the analyst just considers GHGEs emission and nutritional content. Thus, the diet recommendation might be nutritional and low in GHGEs emission yet really expensive for DM. Consequently, the diet recommendation will not be favourable for DM. If the price is included as a criterion, the diet recommendation will be more realistic.
Johnson *et al.* (2014) indicated that humans' diet is affected by various inter-related factors. *Including all criteria* might increase the validity of results. However, it entails a high amount of data and high cognitive load on respondents. It is not always feasible for the analyst to find all the relevant criteria. If the analyst wants to consider all criteria or factors and has a large data set of consumers purchase/consumption history, then machine learning (i.e. collaborative filtering and PageRank) methods can be utilised. The analyst does not have to define criteria or criteria levels because the algorithm will automatically learn the latent factors (criteria) from the collected data (i.e. collaborative filtering). It will not explicitly reveal the weights or part-worth utility, but it will provide diet recommendation for the users.

- **The type of data collection**
*Interactive methods* involve respondents in several stages of examination. It would request a high and intense cognitive burden on respondents. An interactive approach is followed by both MCDA methods and conjoint analysis (i.e. compositional methods and rating/choice-based decompositional methods).
In AHP, the DM(s) is (are) involved from defining the weights of each criterion until the weight of each criterion within one alternative. So, they are involved from the defining the criteria for the evaluation of alternatives. In MACBETH, the DMs are involved from defining the descriptors of the criteria, weighting the criteria, evaluating the alternatives, and also checking the consistency of the weights. The same also applies to bisection method. In order to create a value function, the DM has to answer several iterative questions for each criterion. Due to this high involvement and time required, these methods might be more suitable for cases that involve only a limited amount of respondents. The type of respondents is one of the key issues. Mostly, experts are chosen to

increase the validity of answers when only limited respondents are involved. In conjoint analysis, a focus group should be conducted to specify the representative criteria. Afterwards, the DM has to do the choice experiment to choose the preferable alternatives.

Yet, since these methods do not include actual observation data, there is a possibility that the respondents might think that certain criteria are important, but they are inconsistent in practice. For example, a DM might consider GHGE reduction is more important than price, while in reality she/he always buys the cheapest food/ drink items without considering its sustainability aspect. This problem is referred to as *hypothetical bias*. Conjoint analysis sometimes introduces *a cheap talk* to cope with that (Van Wezemael *et al.*, 2014). The objective is to reduce the chance of respondents giving one opinion but act differently in reality. Thus, respondents are reminded to provide a genuine preference. Even though this bias was clearly stated only in an article about CBCA, this issue could also be applicable in other interactive methods which also do not use actual history data.

An *empirical approach* emphasizes on the observations of actual behavior. For this reason, decompositional methods (i.e. machine learning) could be the suitable methods. Machine learning mainly depends on history data of food purchase or food that has been eaten (food journal). For instance in collaborative filtering, the algorithm can provide suggestion to a user (DM) based on the collected preferences information from many users. So, it is more applicable for *a large amount* of respondents. This method does not always need direct rating from the respondent. It can reduce the cognitive bias, and hypothetical bias from the decision maker since the judgement from the survey is usually prone to that problem (Riabacke et al., 2012). The main problem in machine learning is to obtain a large amount of data, thus it is particularly suitable for data-abundant situation. In some cases, explicit feedback (i.e. direct rating from the user) is used to complement the implicit feedback (i.e. purchase history). Another approach is to use users' profile information (i.e. geographical information) to derive user similarity (Adomavicius and Tuzhilin, 2005).

**-The aggregation level of consumer**

The main consideration is which optimization unit used by the analyst: **average-individuals** within the target group or **an individual (or several individuals)** which represent(s) the respective group. If the elicitation is based on the individual(s), one could consider using the interactive methods (i.e. MCDA and conjoint analysis methods). These methods highly involve DM during the elicitation process. For the **average-individuals approach,** the analyst should consider the preference of all consumers in a target group; thus, the machine learning methods (i.e. CF and PageRank) could be more convenient since they do not require high involvement of DM. They could derive the pattern from purchase or consumption history of DMs.

-**Dealing with uncertain data**

The concern might arise when collecting input for criteria due to the uncertainty of data. For example, the price of food might differ depending on where the food is sold. The analyst should decide whether the variation will be taken into account or not because some target groups might be price sensitive. So, the changes in price might affect their preference towards certain diet. In addition, there are numerous types of uncertainty that can be observed, such as GHGE data. The analyst could hardly find GHGE data for all food or drink items. Thus, one should make estimation. According to Faramitha (2016), there are two main approaches in dealing with this issue, deterministic or stochastics approach. In deterministic approach, there is no uncertainty considered. Meanwhile, the stochastic approach considers randomness or uncertainty. The deterministic approach is most common in MCDA methods (i.e. compositional methods). However, interval SMARTS might be suitable for stochastic approach since the criteria can be modelled as intervals to cover the possible differences between the current/known value and the reality (Mustajoki *et al.*, 2005b). The intervals of value can then be used for price and GHGE data. Decompositional methods can integrate uncertainty into the model from various ways. In conjoint analysis, the uncertainty of data can be incorporated by using a categorical scale. For instance, the level of GHGEs reduction can be divided into two categories: 'high' (25-505) and 'low' (0-25%). However, it is suggested to use a concise level to avoid confusion of respondents (Orme, 2002). The collaborative filtering (i.e. latent factor models) handles this uncertainty by taking into account the 'variety of time' in the algorithm (Koren and Bell, 2011).

# CHAPTERV. CONCLUSION, LIMITATION, AND RECOMMENDATIONS

## *5.1 Conclusion*

Preference elicitation method is useful to capture the preference of individual towards certain criteria of diet modelling. The inclusion of individuals' preferences into the SHARP diet model could improve the practicability of the resulting diet recommendation. There are many preference elicitation methods available and recently applied in various applications. In this section, we summarized the main findings of this literature review and answered the research questions defined in Chapter 1.

***SQ 1: What are the steps/ procedure of those preference elicitation methods?***
A number of preference elicitation methods were reviewed following a snowball approach (see Chapter 3). We classified the preference elicitation methods to compositional and decompositional categories as described in Section 2.1.4. The brief procedure of each method is summarized in Table 9 (*compositional*) and 10 (*decompositional*). Compositional are methods that evaluate the utility of criteria, which are then composed to compute the utility of each alternative. In general, compositional methods ask DM to assess the criteria by *direct rating (i.e. scales, point allocation, SMART), ranking (i.e. direct ranking, SMARTER), pairwise comparison (i.e. AHP, MACBETH), swing weighting (i.e. SMARTS), and scoring function (i.e. bisection)*. Then the values of criteria are then added to calculate the total utility of alternatives. Decompositional are methods that evaluate the utility of alternative(s) then decompose it to obtain the utility of criteria. Firstly, the decompositional methods assess the alternatives using several approaches: *choice-based (PAPRIKA, BWS, CBCA), rating-based (RBCA), and learning-based* (collaborative filtering, PageRank). The utility of criteria or criteria levels can be derived from the results using statistical analysis or linear programming, except for learning-based methods. Instead of providing the utility of criteria or criteria levels, the learning algorithm would capture the patterns of users' consumption and generate a suggestion for a diet alternative. There could be several variations of the method under a category. The modifications of procedure might be simple, but it could have an influence on the cognitive process of the DM.

***SQ 2: What are the main differences and features of existing preference elicitation methods?***
Each method has its own feature that might differentiate it from another method. The summary of the features is presented in Table 33. There is three main components: *(1) Extraction: how the information is derived from respondents/ decision makers (DM); (2) Representation: the format of representing DMs' inputs; (3) Interpretation: how to assign meaning to the obtained information.* The common extraction methods are rating and ranking, while the typical utility model for interpretation is an additive function. Point estimates and ordinal are frequently used to represent DM's input. The semantic descriptive representation is common in pairwise comparison methods (i.e. AHP, MACBETH).

***SQ 3: What are the advantages and disadvantages of each method?***
There are several important parameters to analyse the advantages and disadvantages of each preference elicitation methods. The parameters are *weight typology, uncertainty treatment, robustness, ease of use, sample size, learning dimension, and context of use*. The short descriptions of those parameters can be found in Table 2. The advantages and disadvantages are based on rational evaluation parameters in Table 7. From that information, the parameters can be categorized as advantages or disadvantages. The full list of advantages and disadvantages can be found in Table 35. There are several interesting findings from the results:

- **Sample size**. Machine learning technique might require more data than other methods to provide reliable recommendations compared to other methods. However, there is no exact rule to determine the sample size for almost all the methods.
- **Ease of use**. Most of the methods are easy when the number of criteria and criteria levels are limited, but become cognitively burdensome when they are increased, especially for techniques under pairwise comparison, scoring function, and conjoint analysis.
- **Rank reversal**. Only pairwise comparison techniques are prone to rank reversal because they do not consider all criteria at the same time and transitivity. AHP is more prone to rank reversal than MACBETH.
- **Learning dimension**. The multi-stages methods (i.e. conjoint analysis and pairwise comparison) are not easy for re-evaluation.
- **Dealing with uncertain data**. Most methods have treatments for uncertainty. The imprecise judgement is usually coped with using a value interval, categorical scale, or 'no-buy'/indifferent option. Missing value can also be handled by various techniques, but it might be more difficult for decompositional methods.
- **The context of use**. Knowledge level of the DM should be aligned with the context of elicitation. MCDA and machine learning methods commonly are used to screen product or service, while conjoint analysis methods are designed to determined users' needs.

***SQ 4: What are the current trends in preference elicitation methods?***

The outcome of the systematic search was used to map the trends of preference elicitation methods. The trends show that the usage of decompositional methods exceeds compositional methods significantly from 2001 to 2016. The mostly used approach is conjoint analysis.

***SQ 5: When the methods can be appropriately applied in diet modelling?***

All the obtained methods basically are potentially useful in SHARP diet model application. However, there is a certain condition when particular method(s) is/are more applicable. There are some factors to be considered as provided below.

- *The identification of performance indicators for SHARP diet*
  If only limited number of criteria from SHARP dimension to be considered, then MCDA and conjoint analysis methods (i.e. compositional methods and rating-based/choice-based decompositional) are applicable. If more criteria to be considered and they do not need to be defined, then machine learning methods are appropriate.
- *The type of data collection*
  In a data-scarce situation where information on consumption patterns of individuals is not available, it might be interesting to use MCDA and conjoint analysis because they use the interactive technique to recover weights. In a data-abundant situation where the analyst has sufficient data on consumption patterns of individuals (i.e. actual purchase or consumption history), then machine learning methods which use empirical approach (i.e. collaborative filtering, PageRank) might be more appropriate to use.
- *The aggregation level of consumer*
  If the analyst focuses on average-individuals, then non-interactive or empirical approach might be more convenient. If the focus is on the representative individual(s), the interactive methods (i.e. compositional methods and rating-based/choice-based conjoint analysis) might be more suitable.
- *Dealing with uncertain data*
  If the deterministic approach is used, then MCDA methods (i.e. compositional methods) might be applicable. However, interval SMARTS might be suitable for stochastic approach since the criteria can be modelled as intervals. If the stochastic approach is utilized, then decompositional methods (i.e. conjoint analysis and machine learning) can be selected.

*Main research question:* Based on the literature review, a number of representative preference elicitation methods were obtained. They were *scales, point allocation, ranking, interval SMART, SMART, SMARTER, AHP, MACBETH, bisection, PAPRIKA, BWS, CBCA, RBCA, collaborative filtering, and PageRank*. They could be further classified into compositional and decompositional methods based on their procedures. In general, they could be applicable for diet modelling, depending on the conditions.

The conditions are: (1) the identification of indicators for SHARP diet: *limited or all*; (2) the type of data collection*: interactive or empirical*; (3) the aggregation level of consumer group: *individual or average-individuals;* (4) dealing with uncertain data: *deterministic or stochastic*. In general, machine learning techniques (i.e. CF and PageRank) are more convenient for using all SHARP diet indicators, empirical data collection, the average-individuals consumer level of aggregation, and stochastic approach. While MCDA methods (i.e. compositional methods) and conjoint analysis (i.e. rating-based and choice-based decompositional methods) are more appropriate for limited SHARP diet indicators, interactive data collection, and individual(s) level of aggregation. In terms of dealing with data uncertainty, MCDA methods might use deterministic approach, while conjoint analysis methods use stochastic approach.

### *5.2 Limitations*

This literature review provides a comprehensive overview of the main preference elicitation methods. We used a structured way to identify and review key articles in this domain. Of course, there are some limitations that could be considered for future research. The limitations of this research are:

- This present research focused on giving simple and theoretical examples of preference elicitation methods. These examples could provide a basic understanding of how the methods should be conducted and an indication of their differences. However, they were not able to provide actual utility variations derived from each method and compare one another. This would help to deliver better and more practical suggestions on which preference elicitation methods that could be used particularly for SHARP diet model. Nevertheless, the actual implementation would be a time-consuming process as well. Thus it could not be done in this present research due to the limitation of time.
- The important parameters to compare advantages and disadvantages of preference elicitation methods were mainly obtained from the publication which focused on comparing AHP and Conjoint Analysis. Thus, there might be parameters which were hard to define for other methods, such as ‘context of use’ and ‘missing values.' The common ‘context of use’ of AHP and conjoint analysis methods were already stated in the publication, but not for other methods. The common ‘context of use’ of other methods was derived based on the reviewed publications. It should be cautioned whether the number of publications was sufficient to derive a conclusion, particularly for *ranking, scoring function, direct rating, interval SMART, and PAPRIKA* because the number of reviewed publication was limited. The ‘missing value’ parameter was scarcely mentioned in other publications. To cope with this problem, the group was represented by several methods. For example, the difficulty to handle missing value in MCDA (compositional methods) was represented by AHP and MACBETH. Since the other methods (i.e. ranking and rating) were also simple and not requiring statistical analysis, most likely they also could handle missing value easily. This ‘missing value’ problem might exist in the preference elicitation methods, but probably not too often since electronic/ online survey could help to assure that DM fully filled in the survey. Due to the limitation of time, we used these parameters. Despite this issue, the obtained important parameters were able to provide understanding and comparison of the methods’ advantages and disadvantages.
- Since there was no previous research about selecting preference elicitation methods for diet models, the conditions for applying preference elicitation methods were mostly based on the

conceptual decisions of Faramitha (2016) who performed a review of a mathematical diet modelling approach. Thus, there might be other factors/conditions that were not covered in this research. Even though they might not be complete, the obtained conditions were able to deliver an idea when the preference elicitation methods could be applicable in the SHARP diet.
- Several conference papers which were found during the systematic search did not have author and publication titles. Hence, it reduced the amount of collected publications for collected relevant publications. However, the publications were mainly about decompositional method (i.e. machine learning), and the amount was not too many. Thus, it did not influence the trends of preference elicitation significantly. The result of the trends is still valid.
- Due to the limited background in artificial intelligence and limited time, the machine learning methods were chosen based on the suggestion of supervisors. Thus, there might be other well-known machine learning methods could be found in the literature. The reviewed methods were selected because they were widely used (i.e. in google search, Amazon, and Netflix) and quite easy to understand without detailed technical explanations. At least, they could represent how machine learning could be used as preference elicitation methods.

### *5.3 Recommendations*

Based on the research, there are several factors that could be considered by researcher or policy maker in the usage of preference elicitation methods for the development of SHARP diet model:

*Managerial factors*

- **Knowledge of the respondents**. The identification of SHARP diet indicators plays a great role. If only a limited number of criteria to be considered, the policy maker should assure that the indicators are relevant. Therefore, the involvement of experts (in diet-related field) for the identification of indicators can be helpful.

*Technological factors*

- **The availability of software support and data**. Different preference elicitation might require different support and data type. Thus, the availability of software or computer support and the methods should be aligned. For example, if the data about population's consumption history is abundant and the policy maker has a support for machine learning, then the machine learning methods could be used. Since this method is based on actual behaviour of consumer, the usage of machine learning method (i.e. collaborative filtering) might reduce hypothetical bias and increase the reliability of the optimized diet outcome.
- **Type of food***.* The policy maker or researcher should be able to define the food items that are relevant for European consumer and ensure that the diet recommendation does not deviate too much from that; otherwise, the SHARP diet recommendation will not be applicable.

To follow-up, the next research could focus on confirming the results of this research by conducting actual implementation of diet models with the preference elicitation technique. It would be interesting to observe how actually different preference elicitation methods affect the outcome of diet optimization. The quality of the outcome for each method can be defined by comparing to the actual choices of the individual. The preference elicitation method which could create a more sensible diet recommendation could be then suggested to policy maker or researcher.
In this present research, a snowball method was conducted from 5 leading publications. In order to widen the scope of methods and information, the future research might employ additional leading publications or try with other leading publications. Then, it might be interesting to compare the results with the outcome of present research.

# CHAPTER VI. EVALUATION OF THE RESEARCH

## *6.1 Evaluation of the research process*

The development of conceptual review framework was a critical issue. The list of important parameters and factors of choosing the appropriate method for SHARP diet model are supposed to be specified before conducting the literature review. Nonetheless, the list of important parameters was still updated while conducting the literature review. Additionally, the factors of choosing the appropriate method for SHARP diet were also still being determined when making discussion part. The changes necessitated some adjustments in Chapter 2 as well, in which the concept of diet model and MP were introduced. Thus, it would have been more structured if those works were done before conducting the literature review.

There were three major steps in this research: (1) collecting and classification of methods; (2) understanding the methods (advantages disadvantages, main components, and illustrative examples); (3) evaluation. The first and the second stages were quite time-consuming because most of the methods were not well-known and somewhat technical. Thus, it took more time to understand and put that into the context of diet. Another problem also relates to the structure of the repots to it understandable for readers. Restructuring processes were done several times by using the input from supervisors and other readers.

The keywords for systematic literature review were also altered a few times in order to obtain more representative results. The previous keyword resulted in too few publications, thus in some years, there was no publication observed. Therefore, the search terms for systematic search had to be adjusted.

## *6.2 Reflection of the researcher*

This section is dedicated to reflecting on my role and performance as a researcher during the whole research process. The research topic that I have worked on was challenging and interesting because it could give me a really broad overview, not only of preference elicitation methods but also about mathematical programming for diet models. I learned about mathematical programming before, but not specifically for diet model. Thus, it provided more insight on the versatility of mathematical programming. However, I was not too familiar with preference elicitation methods, particularly about conjoint analysis and machine learning methods. It took quite a lot of time for me to understand those methods since I also did not have a strong mathematical or consumer science background. This research certainly widened my knowledge and deepened my understanding of diet modelling.

Conducting a literature research was a new experience for me as well. I used to deal with only laboratory and trial works which involved quantitative data most of the times. I still have to improve on how to conduct a proper literature review research. The challenging part was Finding the appropriate publications and understanding the technical explanation in the paper. To overcome the latter issue, I usually used other sources to learn about the methods (i.e. books, video). Since this report contains a significant amount of information, another major problem that I encountered was to structure the report. For this issue, I tried to integrate the input from my supervisor and read the reports several times.

In the end, I had to discuss which method should be appropriate for SHARP diet model and in which conditions. Personally, I was not too confident to write about it because this research was based on literature review, not on the real implementation of diet model. However, my supervisor provided me some insights about the feasible way to discuss it. It might be not perfect, but hopefully, this research would give an overview for people who might conduct a preference elicitation process prior to mathematical programming. Reflecting on the journey that I have been through, I am thankful for the knowledge and valuable experiences that I have gained.

## REFERENCES


ADOMAVICIUS, G. and TUZHILIN, A. 2005. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering* **17**, 734-749.

AHN, B.S. 2011. Compatible weighting method with rank order centroid: Maximum entropy ordered weighted averaging approach. *Eur. J. Oper. Res.* **212**, 552-559.

AL-JANABI, H., FLYNN, T.N. and COAST, J. 2011. Estimation of a preference-based carer experience scale. *Medical Decision Making* **31**, 458-468.

ALI, S. and RONALDSON, S. 2012. Ordinal preference elicitation methods in health economics and health services research: using discrete choice experiments and ranking methods. *British medical bulletin*, lds020.

ALOYSIUS, J.A., DAVIS, F.D., WILSON, D.D., TAYLOR, A.R. and KOTTEMANN, J.E. 2006. User acceptance of multi-criteria decision support systems: The impact of preference elicitation techniques. *Eur. J. Oper. Res.* **169**, 273-285.

ALRIKSSON, S. and ÖBERG, T. 2008. Conjoint analysis for environmental evaluation. *Environmental Science and Pollution Research* **15**, 244-257.

ANANDA, J. and HERATH, G. 2009. A critical review of multi-criteria decision making methods with special reference to forest management and planning. *Ecological economics* **68**, 2535-2548.

ANNUNZIATA, A. and VECCHIO, R. 2013. Consumer perception of functional foods: A conjoint analysis with probiotics. *Food Quality and Preference* **28**, 348-355.

ASIOLI, D., NÆS, T., ØVRUM, A. and ALMLI, V. 2016. Comparison of rating-based and choice-based conjoint analysis models. A case study based on preferences for iced coffee in Norway. *Food Quality and Preference* **48**, 174-184.

BALTUSSEN, R., TEN ASBROEK, A., KOOLMAN, X., SHRESTHA, N., BHATTARAI, P. and NIESSEN, L.W. 2007. Priority setting using multiple criteria: should a lung health programme be implemented in Nepal? *Health Policy and Planning* **22**, 178-185.

BANA E COSTA, C.A., CORTE, J.M. and VANSNICK, J.C. 2011. MACBETH (measuring attractiveness by a categorical based evaluation technique). *Wiley Encyclopedia of Operations Research and Management Science*.

BELTON, V. and STEWART, T. 2002. *Multiple criteria decision analysis: an integrated approach*. Springer Science & Business Media.

BOTTOMLEY, P.A. and DOYLE, J.R. 2001. A comparison of three weight elicitation methods: good, better, and best. *Omega* **29**, 553-560.

BOTTOMLEY, P.A., DOYLE, J.R. and GREEN, R.H. 2000. Testing the reliability of weight elicitation methods: direct rating versus point allocation. *Journal of Marketing Research* **37**, 508-513.

BRIEND, A., DARMON, N., FERGUSON, E. and ERHARDT, J.G. 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.

BUTTRISS, J., BRIEND, A., DARMON, N., FERGUSON, E., 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.

CAPUTO, V., NAYGA, R.M. and SCARPA, R. 2013. Food miles or carbon emissions? Exploring labelling preference for food transport footprint with a stated choice study. *Australian Journal of Agricultural and Resource Economics* **57**, 465-482.

CARSON, R.T. and LOUVIERE, J.J. 2011. A Common Nomenclature for Stated Preference Elicitation Approaches. *Environ. Resour. Econ.* **49**, 539-559.

CHANKONG, V. and HAIMES, Y.Y. 2008. *Multiobjective decision making: theory and methodology*. Courier Dover Publications.

CHEUNG, K.L., WIJNEN, B.F., HOLLIN, I.L., JANSSEN, E.M., BRIDGES, J.F., EVERS, S.M. and HILIGSMANN, M. 2016. Using Best–Worst Scaling to Investigate Preferences in Health Care. *PharmacoEconomics* **34**, 1195-1209.

CINELLI, M., COLES, S.R. and KIRWAN, K. 2014. Analysis of the potentials of multi criteria decision analysis methods to conduct sustainability assessment. *Ecol. Indic.* **46**, 138-148.

CLARET, A., GUERRERO, L., AGUIRRE, E., RINCÓN, L., HERNÁNDEZ, M.D., MARTÍNEZ, I., PELETEIRO, J.B., GRAU, A. and RODRÍGUEZ-RODRÍGUEZ, C. 2012. Consumer preferences for sea fish using conjoint analysis: Exploratory study of the importance of country of origin, obtaining method, storage conditions and purchasing price. *Food Quality and Preference* **26**, 259-266.

CLEEMPUT, I., DEVRIESE, S., KOHN, L., DEVOS, C., VAN TIL, J., GROOTHUIS, K., VANDEKERCKHOVE, P. and VAN DE VOORDE, C. 2014. Incorporating societal preferences in reimbursement decisions. Relative importance of decision criteria according to Belgian citizens. *Health Service Research (HSR)*.

CUNNINGHAM, C.E., DEAL, K. and CHEN, Y. 2010. Adaptive choice-based conjoint analysis. *The Patient: Patient-Centered Outcomes Research* **3**, 257-273.

DANIELSON, M. and EKENBERG, L. 2016. A robustness study of state-of-the-art surrogate weights for MCDM. *Group Decision and Negotiation*, 1-15.

DAY, W., AUDSLEY, E. and FROST, A. 2008. An engineering approach to modelling, decision support and control for sustainable systems. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **363**, 527-541.

DE AMO, S., DIALLO, M.S., DIOP, C.T., GIACOMETTI, A., LI, D. and SOULET, A. 2015. Contextual preference mining for user profile construction. *Inf. Syst.* **49**, 182-199.

DE BEKKER-GROB, E.W., DONKERS, B., JONKER, M.F. and STOLK, E.A. 2015. Sample size requirements for discrete-choice experiments in healthcare: a practical guide. *The Patient-Patient-Centered Outcomes Research* **8**, 373-384.

DEFECHEREUX, T., PAOLUCCI, F., MIRELMAN, A., YOUNGKONG, S., BOTTEN, G., HAGEN, T.P. and NIESSEN, L.W. 2012. Health care priority setting in Norway a multicriteria decision analysis. *BMC health services research* **12**, 1.

DIABY, V., SANOGO, V. and MOUSSA, K.R. 2016. ELICIT: An alternative imprecise weight elicitation technique for use in multi-criteria decision analysis for healthcare. *Expert review of pharmacoeconomics & outcomes research* **16**, 141-147.

DOLAN, J.G. 2005. Patient priorities in colorectal cancer screening decisions1. *Health Expectations* **8**, 334-344.

DOMINGOS, P. 2012. A few useful things to know about machine learning. *Communications of the ACM* **55**, 78-87.

DOMSHLAK, C., HÜLLERMEIER, E., KACI, S. and PRADE, H. 2011. Preferences in AI: An overview. *Artificial Intelligence* **175**, 1037-1052.

E COSTA, C.A.B., CARNERO, M.C. and OLIVEIRA, M.D. 2012. A multi-criteria model for auditing a Predictive Maintenance Programme. *Eur. J. Oper. Res.* **217**, 381-393.

ELAHI, M., RICCI, F. and RUBENS, N. 2016. A survey of active learning in collaborative filtering recommender systems. *Computer Science Review*.

ENNEKING, U., NEUMANN, C. and HENNEBERG, S. 2007. How important intrinsic and extrinsic product attributes affect purchase decision. *Food Quality and Preference* **18**, 133-138.

FELLI, J.C., NOEL, R.A. and CAVAZZONI, P.A. 2008. A multiattribute model for evaluating the benefit-risk profiles of treatment alternatives. *Medical Decision Making*.

FRANCESCHET, M. 2011. PageRank: Standing on the shoulders of giants. *Communications of the ACM* **54**, 92-101.

FREIBAUER, A., MATHIJS, E., BRUNORI, G., DAMIANOVA, Z., FAROULT, E., GOMIS, J.G., O'BRIEN, L. and TREYER, S. 2011. Sustainable food consumption and production in a resource-constrained world. *The 3rd SCAR (European Commission–Standing Committee on Agricultural Research) Foresight Exercise*.

FÜRNKRANZ, J. and HÜLLERMEIER, E. 2011. *Preference learning*. Springer.

GOETGHEBEUR, M.M., WAGNER, M., KHOURY, H., LEVITT, R.J., ERICKSON, L.J. and RINDRESS, D. 2012. Bridging health technology assessment (HTA) and efficient health care decision making with multicriteria decision analysis (MCDA) applying the EVIDEM framework to medicines appraisal. *Medical decision making* **32**, 376-388.

GOLAN, O. and HANSEN, P. 2012. Which health technologies should be funded? A prioritization framework based explicitly on value for money. *Israel journal of health policy research* **1**, 1.

GRACIA, A. and DE-MAGISTRIS, T. 2013. Preferences for lamb meat: A choice experiment for Spanish consumers. *Meat science* **95**, 396-402.

GREEN, P.E. and SRINIVASAN, V. 1978. Conjoint analysis in consumer research: issues and outlook. *Journal of consumer research* **5**, 103-123.

GUINÉE, J., HEIJUNGS, R., DE KONING, A., VAN, L., GEERKEN, T., VAN HOLDERBEKE, M., VITO, B.J., EDER, P. and DELGADO, L. 2006. Environmental Impact of Products (EIPRO) Analysis of the life cycle environmental impacts related to the final consumption of the EU25.

HÅKANSSON, A. 2015. Has it become increasingly expensive to follow a nutritious diet? Insights from a new price index for nutritious diets in Sweden 1980-2012. *Food & nutrition research* **59**.

HANISCH, D. 2012. *Management Consulting in Family Businesses*. BoD–Books on Demand.

HANSEN, P. and OMBLER, F. 2008. A new method for scoring additive multi-attribute value models using pairwise rankings of alternatives. *Journal of Multi-Criteria Decision Analysis* **15**, 87-107.

HAUSER, J.R. 2007. A note on conjoint analysis. *Massachusetts Institute of Technology, MIT Sloan Courseware*.

HONKANEN, P. and FREWER, L. 2009. Russian consumers' motives for food choice. *Appetite* **52**, 363-371.

HORGAN, G.W., PERRIN, A., WHYBROW, S. and MACDIARMID, J.I. 2016. Achieving dietary recommendations and reducing greenhouse gas emissions: modelling diets to minimise the change from current intakes. *International Journal of Behavioral Nutrition and Physical Activity* **13**, 1.

HUBER, J., WITTINK, D.R., FIEDLER, J.A. and MILLER, R. 1993. The effectiveness of alternative preference elicitation procedures in predicting choice. *Journal of Marketing Research* **30**, 105.

JOHNSON, S.R., NADEN, R.P., FRANSEN, J., VAN DEN HOOGEN, F., POPE, J.E., BARON, M., TYNDALL, A., MATUCCI-CERINIC, M., DENTON, C.P. and DISTLER, O. 2014. Multicriteria decision analysis methods with 1000Minds for developing systemic sclerosis classification criteria. *Journal of clinical epidemiology* **67**, 706-714.

KARNIOUCHINA, E.V., MOORE, W.L., VAN DER RHEE, B. and VERMA, R. 2009. Issues in the use of ratings-based versus choice-based conjoint analysis in operations management research. *Eur. J. Oper. Res.* **197**, 340-348.

KHALILI, N.R. and DUECKER, S. 2013. Application of multi-criteria decision analysis in design of sustainable environmental management system framework. *Journal of Cleaner Production* **47**, 188-198.

KINGSTON-SMITH, A.H., EDWARDS, J.E., HUWS, S.A., KIM, E.J. and ABBERTON, M. 2010. Plant-based strategies towards minimising 'livestock's long shadow'. *Proceedings of the Nutrition Society* **69**, 613-620.

KOREN, Y. and BELL, R. 2011. Advances in collaborative filtering. In *Recommender systems handbook* pp. 145-186, Springer.

KOTSIANTIS, S.B., ZAHARAKIS, I. and PINTELAS, P. 2007. Supervised machine learning: A review of classification techniques. In.

LLOYD, A.J. 2003. Threats to the estimation of benefit: are preference elicitation methods accurate? *Health economics* **12**, 393-402.

LOUVIERE, J.J., FLYNN, T.N. and CARSON, R.T. 2010. Discrete choice experiments are not conjoint analysis. *Journal of Choice Modelling* **3**, 57-72.

MAILLOT, M., VIEUX, F., AMIOT, M.J. and DARMON, N. 2010. Individual diet modeling translates nutrient recommendations into realistic and individual-specific food choices. *The American journal of clinical nutrition* **91**, 421-430.

MALCZEWSKI, J. 1999. *GIS and multicriteria decision analysis*. John Wiley & Sons.

MARSH, K., IJZERMAN, M., THOKALA, P., BALTUSSEN, R., BOYSEN, M., KALÓ, Z., LÖNNGREN, T., MUSSEN, F., PEACOCK, S. and WATKINS, J. 2016. Multiple Criteria Decision Analysis for Health Care Decision Making—Emerging Good Practices: Report 2 of the ISPOR MCDA Emerging Good Practices Task Force. *Value in health* **19**, 125-137.

METZGAR, M., RIDEOUT, T.C., FONTES-VILLALBA, M. and KUIPERS, R.S. 2011. The feasibility of a Paleolithic diet for low-income consumers. *Nutrition research* **31**, 444-451.

MILES, M.B. and HUBERMAN, A.M. 1994. *Qualitative data analysis: An expanded sourcebook*. Sage.

MILLWARD, D.J. and GARNETT, T. 2010. Plenary Lecture 3 Food and the planet: nutritional dilemmas of greenhouse gas emission reductions through reduced intakes of meat and dairy foods. *Proceedings of the nutrition society* **69**, 103-118.

MÜHLBACHER, A.C., KACZYNSKI, A., ZWEIFEL, P. and JOHNSON, F.R. 2016. Experimental measurement of preferences in health and healthcare using best-worst scaling: an overview. *Health economics review* **6**, 1.

MUSTAJOKI, J., HÄMÄLÄINEN, R.P. and SALO, A. 2005a. Decision support by interval SMART/SWING-Methods to incorporate uncertainty into multiattribute analysis. *Decision Sciences* **36**, 317-339.

MUSTAJOKI, J., HÄMÄLÄINEN, R.P. and SALO, A. 2005b. Decision support by interval SMART/SWING—incorporating imprecision in the SMART and SWING methods. *Decision Sciences* **36**, 317-339.

NÆS, T., BROCKHOFF, P. and TOMIC, O. 2011. *Statistics for sensory and consumer science*. John Wiley & Sons.

NIKOU, S., MEZEI, J. and SARLIN, P. 2015. A process view to evaluate and understand preference elicitation. *Journal of Multi-Criteria Decision Analysis* **22**, 305-329.

OLIVEIRA, M.D., RODRIGUES, T.C., E COSTA, C.A.B. and DE SÁ, A.B. 2012. Prioritizing health care interventions: A multicriteria resource allocation model to inform the choice of community care programmes. In *Advanced decision making methods applied to health care* pp. 141-154, Springer.

ORME, B. 2002. Formulating attributes and levels in conjoint analysis. *Sawtooth Software Research Paper*, 1-4.

PERIGNON, M., MASSET, G., FERRARI, G., BARRÉ, T., VIEUX, F., MAILLOT, M., AMIOT, M.-J. and DARMON, N. 2016. How low can dietary greenhouse gas emissions be reduced without impairing nutritional adequacy, affordability and acceptability of the diet? A modelling study to guide sustainable food choices. *Public health nutrition*, 1-13.

POLATIDIS, H., HARALAMBOPOULOS, D.A., MUNDA, G. and VREEKER, R. 2006. Selecting an appropriate multi-criteria decision analysis technique for renewable energy planning. *Energy Sources, Part B* **1**, 181-193.

PÖYHÖNEN, M. and HÄMÄLÄINEN, R.P. 2001. On the convergence of multiattribute weighting methods. *Eur. J. Oper. Res.* **129**, 569-585.

RAFFENSPERGER, J.F. 2008. The least-cost low-carbohydrate diet is expensive. *Nutrition research* **28**, 6-12.

RIABACKE, M., DANIELSON, M. and EKENBERG, L. 2012. State-of-the-art prescriptive criteria weight elicitation. *Advances in Decision Sciences* **2012**.

RIBAL, J., FENOLLOSA, M.L., GARCÍA-SEGOVIA, P., CLEMENTE, G., ESCOBAR, N. and SANJUÁN, N. 2016. Designing healthy, climate friendly and affordable school lunches. *The International Journal of Life Cycle Assessment* **21**, 631-645.

SAITO, H. and SAITO, Y. 2013. Motivations for Local Food Demand by Japanese Consumers: A Conjoint Analysis with Reference-Point Effects. *Agribusiness* **29**, 147-161.

SJÖRS, C., RAPOSO, S.E., SJÖLANDER, A., BÄLTER, O., HEDENUS, F. and BÄLTER, K. 2016. Diet-related greenhouse gas emissions assessed by a food frequency questionnaire and validated using 7-day weighed food records. *Environmental Health* **15**, 1.

SWANCUTT, D.R., GREENFIELD, S.M. and WILSON, S. 2008. Women's colposcopy experience and preferences: a mixed methods study. *BMC women's health* **8**, 1.

THOKALA, P., DEVLIN, N., MARSH, K., BALTUSSEN, R., BOYSEN, M., KALO, Z., LONGRENN, T., MUSSEN, F., PEACOCK, S. and WATKINS, J. 2016. Multiple Criteria Decision Analysis for Health Care Decision Making—An Introduction: Report 1 of the ISPOR MCDA Emerging Good Practices Task Force. *Value in health* **19**, 1-13.

TORRANCE, G.W., FEENY, D. and FURLONG, W. 2001. Visual Analog Scales Do They Have a Role in the Measurement of Preferences for Health States? *Medical Decision Making* **21**, 329-334.

VAN DE KAA, E.J. 2010. Prospect theory and choice behaviour strategies: review and synthesis of concepts from social and transport sciences. *European Journal of Transport and Infrastructure Research (EJTIR), 10 (4), 2010*.

VAN WEZEMAEL, L., CAPUTO, V., NAYGA, R.M., CHRYSSOCHOIDIS, G. and VERBEKE, W. 2014. European consumer preferences for beef with nutrition and health claims: A multi-country investigation using discrete choice experiments. *Food Policy* **44**, 167-176.

WANG, J.-J., JING, Y.-Y., ZHANG, C.-F. and ZHAO, J.-H. 2009. Review on multi-criteria decision analysis aid in sustainable energy decision-making. *Renewable and Sustainable Energy Reviews* **13**, 2263-2278.

WHITEHEAD, J.C., PATTANAYAK, S.K., VAN HOUTVEN, G.L. and GELSO, B.R. 2008. Combining revealed and stated preference data to estimate the nonmarket value of ecological services: an assessment of the state of the science. *Journal of Economic Surveys* **22**, 872-908.

YUSOP, Z.B. 2015. Noorul Hassan Zardari Kamal Ahmed Sharif Moniruzzaman Shirazi.

# APPENDICES

## *Appendix I. The list of literature for systematic search*

| | Authors (year) | Title | Methods | Category | Scopus | Web of Science |
|---|---|---|---|---|---|---|
| 1 | Holm et al., 2016 | Enhancing Agent-Based Models with Discrete Choice Experiments | CBCA/DCE | Decompositional | | v |
| 2 | van Dijk et al., 2016 | An Empirical Comparison of Discrete Choice Experiment and Best-Worst Scaling to Estimate Stakeholders' Risk Tolerance for Hip Replacement Surgery | CBCA/DCE,BWS | Decompositional; Decompositional | | v |
| 3 | Morel et al., 2016 | Quantifying benefit-risk preferences for new medicines in rare disease patients and caregivers | CBCA/DCE | Decompositional | | v |
| 4 | Huang et al., 2016 | Consumer Preference Elicitation of Complex Products Using Fuzzy Support Vector Machine Active Learning | Collaborative filtering | Decompositional | v | |
| 5 | Robinson et al., 2015 | A Framework for Estimating Health State Utility Values within a Discrete Choice Experiment: Modeling Risky Choices | DCE | Decompositional | | v |
| 6 | Hollin et al., 2015 | Caregiver Preferences for Emerging Duchenne Muscular Dystrophy Treatments: A Comparison of Best-Worst Scaling and Conjoint Analysis | BWS, CBCA/CDE | Decompositional; Decompositional | | v |
| 7 | Abraham et al., 2015 | Complex antithrombotic therapy: Determinants of patient preference and impact on medication adherence | ACA | Decompositional | v | v |
| 8 | Hess et al., 2015 | Preference Elicitation Tool for Abnormal Uterine Bleeding Treatment: A Randomized Controlled Trial | ACA | Decompositional | v | v |
| 9 | Sundarraj et al., 2015 | On integrating an IS success model and multicriteria preference analysis into a system for cloud-computing investment decisions | AHP-DEA | Compositional | v | |
| 10 | Nikou et al., 2015 | A Process View to Evaluate and Understand Preference Elicitation | AHP, CA | Compositional, Decompositional | v | |
| 11 | Hopfgartner, 2015 | Join the living lab: Evaluating news recommendations in real-time | Recommender system | Decompositional | v | |
| 12 | Vetschera et al., 2014 | Implausible alternatives in eliciting multi-attribute value functions | RBCA | Decompositional | v | |
| 13 | Yang et al., 2014 | Decision support for preference elicitation in multi-attribute electronic procurement auctions through an agent-based intermediary | Pairwise comparison of alternatives (holistic preference) | Decompositional | v | |
| 14 | Rotter, 2014 | Relevance feedback based on n-tuplewise comparison and the ELECTRE methodology and an application in content-based image retrieval | Content-based Image Retrieval | Decompositional, | v | v |
| 15 | Kirchhoff et al., 2014 | A conjoint analysis framework for evaluating user preferences in machine translation | CBCA/CDE | Decompositional | v | |
| 16 | Hollin et al., 2014 | Caregiver Preferences for Emerging Duchenne Muscular Dystrophy Treatments: A Comparison of Best-Worst Scaling and Conjoint Analysis | CBCA/CDE, BWS | Decompositional, Decompositional | v | v |
| 17 | Lee et al., 2013 | Homeowners' decision-making in a premise plumbing failure-prone area | CA | Decompositional | | v |
| 18 | Kovalsky et al., 2013 | Do Consumers Really Know How Much They Are Willing to Pay? | Rating | Compositional | | v |
| 19 | Najafzadeh et al., 2013 | Genomic testing to determine drug response: Measuring preferences of the public and patients using Discrete Choice Experiment (DCE) | CBCA/DCE | Decompositional | v | |

| No | Authors (year) | Title | Methods | Category | Scopus | Web of Science |
|---|---|---|---|---|---|---|
| 20 | Boesch et al., 2013 | Enhancing Validity and Reliability Through Feedback-Driven Exploration: A Study in the Context of Conjoint Analysis | ACA | Decompositional | v | |
| 21 | Gabillon et al., 2013 | Adaptive submodular maximization in bandit setting | Reinforcement learning | Decompositional | v | |
| 22 | Schönberg et al., 2012 | Simulation-based design of wind-Turbine sealing solutions using a systematic approach to robust concept exploration | CA | Decompositional | v | |
| 23 | Rokach et al., 2012 | Initial profile generation in recommender systems using pairwise comparison | Collaborative filtering | Decompositional | v | |
| 24 | Ren et al., 2011 | A Design Preference Elicitation Query as an Optimization Process | Support Vector Machine | Decompositional | | |
| 25 | Bornemann et al., 2011 | Psychological Distance and the Dual Role of Price | Scale (7-point semantic scale) on product alternatives (RBCA) | Decompositional | v | v |
| 26 | Danner et al., 2011 | Integrating patients' views into health technology assessment: Analytic hierarchy process (AHP) as a method to elicit patient preferences | AHP | Compositional | | v |
| 27 | Michalek et al., 2011 | Enhancing marketing with engineering: Optimal heterogeneous markets | CBCA/DCE | Decompositional | v | v |
| 28 | Rosenthal et al., 2011 | Using decision-theoretic experience sampling to build personalized mobile phone interruption models | Machine learning | Decompositional | v | |
| 29 | Van Houtven et al., 2011 | Eliciting benefit-risk preferences and probability-weighted utility using choice-format conjoint analysis | CBCA/DCE | Decompositional | v | |
| 30 | Pieterse et al., 2010 | Methodologic evaluation of adaptive conjoint analysis to assess patient preferences: An application in oncology | ACA | Decompositional | v | v |
| 31 | Scholz et al., 2010 | Measuring consumer preferences for complex products: A compositional approach based on paired comparisons | Paired comparison-based preference measurement (PCPM), ACA | Decompositional | v | v |
| 32 | Mueller et al., 2010 | What you see may not be what you get: Asking consumers what matters may not reflect what they choose | BWS | Decompositional | | v |
| 33 | Chang et al., 2009 | How Closely Do Hypothetical Surveys and Laboratory Experiments Predict Field Behavior? | CBCA/DCE | Decompositional | | v |
| 34 | Johnson et al., 2009 | Using Conjoint Analysis to Estimate Healthy-Year Equivalents for Acute Conditions: An Application to Vasomotor Symptoms | CBCA/DCE | Decompositional | | v |
| 35 | Chu et al., 2009 | Interactive learning of independent experts' criteria for rescue simulations | Supervised machine learning, CBCA | Decompositional | v | v |
| 36 | Chica et al., 2009 | Integration of an emo-based preference elicitation scheme into a multi-objective aco algorithm for time and space assembly line balancing | Goal programming (preference per criterion) | Compositional | v | |
| 37 | MacDonald et al., 2009 | Preference inconsistency in multidisciplinary design decision making | CBCA/DCE | Decompositional | v | |
| 38 | van Tijl et al., 2008 | The Use of Multi-Criteria Decision Analysis Weight Elicitation Techniques in Patients with Mild Cognitive Impairment A Pilot Study | CA | Decompositional | | v |
| 39 | Braziunas et al., 2008 | Elicitation of factored utilities | GAI models | Decompositional | v | |

| No | Authors (year) | Title | Methods | Category | Scopus | Web of Science |
|---|---|---|---|---|---|---|
| 40 | Dagher et al., 2008 | Elicitation and modeling of customers' preferences in industrial design: A comparative study on vehicle front end | RBCA | Decompositional | v | |
| 41 | Lusk et al., 2008 | An incentive compatible conjoint ranking mechanism | RBCA | Decompositional | v | v |
| 42 | Kaplan et al., 2007 | A new MCDM approach to solve public sector planning problems | Decision rule (with machine learning algorithm) | Decompositional | v | |
| 43 | Sell et al., 2007 | Ecosystem services from tropical forestry projects - The choice of international market actors | Holistic choice, Direct rating | Decompositional, Compositional | v | v |
| 44 | Telser et al., 2007 | Validity of discrete-choice experiments evidence for health risk reduction | CBCA/DCE | Decompositional | | v |
| 45 | Osssa et al., 2007 | Recombinant erythropoietin for chemotherapy-related anemia - Economic value and health-related quality-of-life assessment using direct utility elicitation and discrete choice experiment methods | TTO | Decompositional | | v |
| 46 | Domshlak et al., 2007 | Efficient and non-parametric reasoning over user preferences | Support vector machine | Decompositional | v | v |
| 47 | Myers, 2006 | The impact of promotions on the convergence in preference elicitation methods | Constant sum-scale, conjoint measurement, actual choice, CA | Compositional; Decompositional | v | |
| 48 | Sassi et al., 2005 | Conjoint analysis of preferences for cardiac risk assessment in primary care | CA | Decompositional | v | v |
| 49 | Millet et al., 2005 | Incorporating negative values into the Analytic Hierarchy Process | AHP | Compositional | v | |
| 50 | Wanyama et al., 2005 | Using prediction to provide decision support for the elicitation of user preferences | Neural networks | Decompositional | v | |
| 51 | Avesani et al., 2005 | Collaborative case-based preference elicitation | Collaborative filtering | Decompositional | v | |
| 52 | Gajos et al., 2005 | Preference elicitation for interface optimization | ARNAULD | Decompositional | v | |
| 53 | Avesani et al., 2005 | Facing scalability issues in requirements prioritization with machine learning techniques | Case-based ranking | Decompositional | v | |
| 54 | DeSarbo et al., 2005 | Evolutionary preference/utility functions: A dynamic perspective | Rating-based CA | Decompositional | v | |
| 55 | Lee et al., 2004 | Learning user preferences for wireless services provisioning | | Decompositional | v | |
| 56 | Cameron et al., 2002 | Alternative non-market value-elicitation methods: Are the underlying preferences the same? | CA | Decompositional | v | |
| 57 | Johnson et al., 2001 | Sources and effects of utility-theoretic inconsistency in stated-preference surveys | Rating-based CA | Decompositional | v | |
| 58 | Paik et al., 2001 | Applying natural language processing (NLP) based metadata extraction to automatically acquire user preferences | natural language processing (NLP) | Decompositional | v | |

## Appendix II. The list of literature from snowball methods

| No | Author (year) | Title | Objective |
|---|---|---|---|
| 1 | Goetghebeur et al., 2012 | Bridging Health Technology Assessment (HTA) and Efficient Health Care Decision Making with Multi-criteria Decision Analysis (MCDA) Applying the EVIDEM Framework to Medicines Appraisal | to express individual perspectives of policy and decision makers by assigning weights to each criterion of the MCDA core model |
| 2 | Bottomley and Doyle, 2001 | A comparison of three weight elicitation methods: good, better, and best | to compare DR, Max100, and Min10 |
| 3 | Hein et al., 2008 | Comparison of five common acceptance and preference methods | to evaluate the existing breakfast bars. Familiarity needed. |
| 4 | Kroese et al., 2010 | A Framework for the Prioritization of Investment in the Provision of Genetic Tests | to evaluate genetic tests and rank them for entry onto genetic testing. |
| 5 | Cleemput et al., 2014 | Incorporating societal preferences in reimbursement decisions relative importance of decision criteria according to Belgian citizens | to measure public preference weights for reimbursement criteria |
| 6 | Nikou et al., 2015 | A Process View to Evaluate and Understand Preference Elicitation | to compare AHP and CA |
| 7 | Dolan et al., 2005 | Patient priorities in colorectal cancer screening decisions | to examine how a group of patients established priorities when making trade-offs of the five currently recommended colorectal cancer screening program |
| 8 | Mustajoki et al., 2005 | Decision support by interval SMART/SWING - Incorporating imprecision in the SMART and SWING methods | to evaluate alternatives. No familiarity needed |
| 9 | Oliviera et al., 2012 | Prioritizing health care interventions: A multi-criteria resource allocation model to inform the choice of community care programs | to select and re-design program. No familiarity needed. |
| 10 | Pinhero et al., 2008 | A Multi-criteria Model Applied in the Diagnosis | to diagnose Alzheimer disease |
| 11 | e Costa et al., 2012 | A multi-criteria model for auditing a Predictive Maintenance Program | to evaluate a Predictive Maintenance Program. |
| 12 | European Medicines Agency, 2011 | Benefit-Risk Methodology Project | to develop the testing tools and processes for balancing multiple benefits and risks, as an aid to provide science-based regulatory decisions about medicinal products |
| 13 | Felli et al., 2009 | A multi-attribute model for evaluating the benefit-risk profiles of treatment alternatives. | to evaluate a medicine's pre- and post-marketing performance |
| 14 | Bottomley et al., 2000 | Testing the Reliability of Weight Elicitation Methods: Direct Rating versus Point Allocation | to evaluate car before buying |
| 17 | Pöyhönen and Hämäläinen, 2001 | On the convergence of multi-attribute weighting methods | to re-examine the properties of the methods in a computer-aided study |
| 15 | Tervonen, 2015 | Applying multiple criteria decision analysis to comparative benefit-risk assessment: choosing among statins in primary prevention | to provide guidelines for applying MCDA in benefit risk assessment |
| 17 | Hansen et al., 2012 | A new process for creating points systems for prioritizing patients for elective health services | to prioritize patients for surgery based on point systems from. No familiarity needed. |
| 18 | Golan & Hansen, 2012 | Which health technologies should be funded? A prioritization framework based explicitly on value for money | to evaluate which health technologies should be funded. No familiarity needed. |
| 19 | Johnson et al., 2014 | Multi-criteria decision analysis methods with 1000Minds for developing systemic sclerosis classification criteria | to develop classification criteria for systemic sclerosis (SSc) are being developed |
| 20 | Al Janabi et al., 2011 | Estimation of a Preference-Based Career Experience Scale | to estimate preference-based index values for caring experience |
| 21 | Swancutt et al., 2008 | Women's colposcopy experience and preferences: a mixed | to identify patients' preferences for aspects of appointments within the colposcopy service and to make suggestions for service improvement. |
| 22 | Hein et al., 2008 | Comparison of five common acceptance and preference methods | to compare the above five consumer acceptance and preference test methods |
| 23 | Asioli et al., 2013 | Comparison of rating-based and choice-based conjoint analysis models. A case study based on preferences for iced coffee in Norway | to compare RBCA and CBCA in product evaluations |

| No | Author (year) | Title | Objective |
|---|---|---|---|
| 24 | Almli et al., 2015 | Investigating individual preferences in rating and ranking conjoint experiments. A case study on semi-hard cheese | to compare conjoint measurement and self-explicated method |
| 25 | Gracia and de-Magistris, 2013 | Preferences for lamb meat: A choice experiment for Spanish consumers | to analyze consumers' preferences for different lamb meat attributes |
| 26 | Annunziata and Vecchio, 2012 | Consumer perception of functional foods: A conjoint analysis with probiotics | to explore consumers evaluation of four attributes of probiotics functional foods |
| 27 | Karniouchina et al., 2008 | Issues in the use of ratings-based versus choice-based conjoint analysis in operations management research | to review recent developments and provide new evidence on how the choice of different variants of conjoint analysis might affect study results (RBCA vs CBCA) |
| 28 | Caputo, 2013 | Food miles or carbon emissions? Exploring labelling preference for food transport footprint with a stated choice study | to evaluate which type of information is preferred by consumers in food labels |
| 29 | Wezemeel et al., 2014 | European consumer preferences for beef with nutrition and health claims: A multi-country investigation using discrete choice experiments | to investigate consumer preferences for nutrition labelling on beef steaks |
| 30 | Saito, 2012 | Motivations for Local Food Demand by Japanese Consumers: A Conjoint Analysis with Reference-Point Effects | to identify the main motivations behind the purchase of locally produced foods. |
| 31 | Baltussen et al., 2007 | Priority setting using multiple criteria: should a lung health programme be implemented in Nepal? | To identify and weight the various criteria for priority setting, and to assess the health program |
| 32 | Marsh et al., 2012 | Prioritizing investments in public health: a multi-criteria | Review of methods |
| 33 | Defechereux et al., 2012 | Health care priority setting in Norway a multi-criteria decision analysis | to compare the values of the country's health policy makers with these three official principles. |
| 34 | Elahi et al., 2016 | A survey of active learning in collaborative filtering recommender systems | to review collaborative filtering |
| 35 | Koren & Bill, 2011 | Advances in collaborative filtering | to review collaborative filtering |
| 37 | Desrosiers & Karypis, 2011 | A Comprehensive Survey of Neighborhood-based Recommendation Methods | Review of methods |
| 38 | Adomavicius and Tuzhilin, 2005 | Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions | Review of methods |
| 39 | Franceshet, 2011 | PageRank: Standing on the Shoulder of Giants | to review PageRank method in various applications |
| 40 | Grover, 2012 | Comparative Analysis Of PageRank And HITS Algorithms | to compare PageRank and HITS algorithm |
| 41 | Diaby et al., 2016 | ELICIT: An alternative imprecise weight elicitation technique for use in multi-criteria decision analysis for healthcare | to introduce ELICIT |
| 42 | Danielson and Ekenberg, 2016 | A robustness study of state-of-the-art surrogate weights for MCDM | to compare state-of-the-art surrogate weights |

## *Appendix III. The data analysis for snowball methods*

*Note : the publications are ordered based on the methods and numbered once.*

| *No* | *Author (year)* | *Methods* | *Weight typology* | *Uncertainty treatment* | *Ease of use* | *Sample size* | *Robustness* | *Learning dimension* | *Context of use* |
|---|---|---|---|---|---|---|---|---|---|
| *1* | Goetghebeur et al., 2012 | Scales | Importance | NA | Easy, the scoring process helps respondents think through each issue. Software support : Excel | 13 experts | | | Screen products, familiarity needed |
| *2* | Bottomley and Doyle, 2001 | Scales | Importance | Consistency check: test and retest reliability (more reliable than PA) | | 36 | No rank reversal | One stage (relatively easy to re-evaluate) | Determine user's needs, no familiarity needed |
| *3* | Hein et al., 2008 | Scales | Importance (central tendency) | No re-tasting or review of previous evaluation was permitted after completion of a sample. | Easy, lack of freedom to elicit responses due to defined response categories, individual scale use and can be interpreted differently across cultures | 233 divided to 5 countries | | | Screen existing products, familiarity needed |
| *4* | Kroese et al., 2010 | Point Allocation (100) | Trade-off | | | | | | Screen products (generic test), familiarity needed. |
| | Bottomley and Doyle, 2001 | Point Allocation (100) | Trade-off | Tend to give 50% more weight to their most important attribute compared to those using DR | | 36 | | | |
| | Bottomley and Doyle, 2001 | Scales (Max100) | Importance | Consistency check: test and retest reliability (Max 100 & DR are more reliable than Min 10) | Easy, DR and Max100 are easier than Min 10 | 36 | | | |
| | Bottomley and Doyle, 2001 | Scales (Min10) | Importance | | | 36 | | | |
| *5* | Cleemput et al., 2014 | AHP | Trade-off | Consistency check: Consistency Index (Generally, high levels of consistency reported) | Some authors that education does not limit the understanding of questions. | | Rank reversal | One-stage | Screen products, familiarity needed |
| *6* | Nikou et al., 2015 | AHP | Trade-off | It can deal with missing values easily | It is easy for only limited amount of criteria | Large number required | Rank reversal | Easy to incorporate additional criteria | Screen product or service, familiarity needed |
| | Bottomley and Doyle, 2001 | AHP | Trade-off | Built-in check on the consistency of the | Relative ease of use | 48 | Rank reversal | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | judgments | | | | | |
| 7 | Dolan et al., 2005 | AHP | Trade-off | | | | | | Screen/ screen products, familiarity needed |
| 8 | Mustajoki et al., 2005 | Interval SMART | Importance relative to the reference attribute | - Intervals to the weight ratio questions to describe possible imprecision<br>- An empty feasible region indicates inconsistency | - Cognitive simplicity of SMART/SWING by using interval value.<br>- Able to select reference attribute<br>- Larger problem, more difficult computation<br>-WINPRE software support | NA | NA | Easy to re-evaluate, we can continue by adjusting the other attributes or alternatives until the best alternative is found | Screen products, familiarity needed |
| | Cleemput et al., 2014 | Ranking | Importance. No indication about the true importance, just acknowledge that one is more preferable than others | Internal consistency weak | - Easy to complete<br>- Easy to analyse | NA | No, because all criteria are directly considered | One stage, easy to re-evaluate | NA |
| 9 | Oliviera et al., 2012 | MACBETH | Trade off | Consistency check : DM can validate the weights, sensitivity analysis on differences among DM | - The definition of "base" and "target" improve the intelligibility of the criteria<br>- Avoiding the difficulty of numerical assessment and cognitive uneasiness<br>- M-MACBETH and PROBE software support | NA | Rank reversal | Multi stage ($1^{st}$: defining the descriptors; $2^{nd}$: elicitation) | Screen products, familiarity needed |
| 10 | Pinhero et al., 2008 | MACBETH | Trade off | Consistency check : sensitivity analysis and dominance | A difficult task for analyst, but it forms good judgement<br>HI-VIEW (consistency check), M-MACBETH | 44 | Rank reversal | Multi stage (defining the descriptor first) | Screen disease (diagnose) |
| 11 | e Costa et al., 2012 | MACBETH | Trade off | DM checks the weight assessed | M-MACBETH software support | NA | NA | NA | Screen a program, familiarity needed |
| 12 | European Medicines Agency, 2011 | SWING | Importance | -Extensive sensitivity analysis by checking how the balance changed<br>- DM can check their judgements | Software support with added value-graph | | | | Screen products, experts needed |
| 13 | Felli et al., 2009 | SWING | Trade-off | NA | NA | NA | NA | NA | Screen products, experts needed |
| 14 | Bottomley et al.,2000 | SWING | | | | | | | Determine user's needs, no familiarity needed |

| No | Author (year) | Methods | Weight typology | Uncertainty treatment | Ease of use | Sample size | Robustness | Learning dimension | Context of use |
|---|---|---|---|---|---|---|---|---|---|
| 15 | Pöyhönen and Hämäläinen, 2001 | SWING | Trade-off | | | varied | No, all criteria considered at once | | Determine user's needs, no familiarity needed |
| | Pöyhönen and Hämäläinen, 2001 | SMART | | | | varied | | | |
| | Pöyhönen and Hämäläinen, 2001 | SMARTER | | | | varied | | | |
| | Pöyhönen and Hämäläinen, 2001 | Direct Point Allocation | | | | varied | | | |
| 16 | Tervonen, 2015 | Bisection | | | | | | | Screen product, familiarity needed |
| | Bottomley et al., 2000 | Point Allocation (DR) | | | | 113 students | | | |
| | Bottomley et al., 2000 | Direct rating (DR) | | | | 113 students | | | |
| 17 | Hansen et al., 2012 | PAPRIKA | Trade-off between two criteria | Consistency check: test and retest reliability | Easy to understand | NA | Yes, because all data is stored and software analyse transitivity | | Screen patient prioritizing, familiarity needed |
| 18 | Golan & Hansen, 2012 | PAPRIKA | Trade-off between two criteria | Consistency check: test and retest reliability | - Eliminates all other possible questions that are implicitly answered as corollaries of those already answered (by logical property of "transitivity")<br>- Software support (1000Minds, PAPRIKA) | 61 (mixed of professionals, experts) | No, software analyse transitivity | Easy to re evaluate because all data is stored | Screen product that should be funded, familiarity needed |
| 19 | Johnson et al., 2014 | PAPRIKA | Trade-off between two criteria | Consistency check: test and retest reliability | -Software eliminates some other questions, pairwise comparison is easy, administered through internet<br>- 1000Minds, PAPRIKA | 8 experts | No, software analyse transitivity | Easy to re evaluate because all data is stored | Understand user's needs |
| | Cleemput et al., 2014 | BWS | Trade-off | Consistency check: test-retest reliability, but lower than DCE.<br>OR in most cases – not differ significantly after rescaling | BWS is easier than ranking, but DCE is more preferred than BWS<br>The number of scenario can be reduced by factorial design | | | Multi stage | |
| 20 | Al Janabi et al., 2011 | BWS | Trade-off | NA | NA | 150 (actually min. 30) | | | Screen experience, familiarity needed |

| No | Author (year) | Methods | Weight typology | Uncertainty treatment | Ease of use | Sample size | Robustness | Learning dimension | Context of use |
|---|---|---|---|---|---|---|---|---|---|
| 21 | Swancutt et al., 2008 | BWS | Trade-off | Only fully completed questionnaire will be analysed, thus no extreme scenarioare present | Easier than DCEs because there is rejection option Accompanied by information leaflet | 30 respondents (min. @group for significant result) | | | Screen and improve service, familiarity needed |
| 22 | Hein et al., 2008 | BWS | Trade-off | No re-tasting or review of previous evaluation was permitted after completion of a triad | -Forced choice<br>-No direct rating scale is involved which eliminates lack of freedom<br>- DM is more confident to provide accurate information<br>- It takes longer time to finish than DCE | 233 divided to 5 countries | | Multi-stage | Screen product preference, familiarity needed |
| 23 | Asioli et al., 2013 | RBCA | Trade-off | significant value | Data collection with Eye-Question system | 101 | | | Screen consumer's preference, familiarity needed |
| 24 | Almli et al., 2015 | RBCA | Trade-off | | | | | | Investigate preference on product |
| | Asioli et al., 2013 | RBCA | Trade-off | NA | -Easy to perform by DM ANOVA analysis of rating data was easier to perform than choice-based<br>-Software support MiniTab | 101 | | Multi-stage | |
| 25 | Gracia and de-Magistris, 2013 | RBCA | Trade-off | NA | -No buy option<br>-Pilot survey to test the survey<br>-SPSS software support | 266 (according to confidence interval and error) | | Multi-stage | Screen consumers' preferences, familiarity needed |
| 26 | Annunziata and Vecchio, 2012 | RBCA | Trade-off | NA | NA | 600 | NA | Multi-stage | Determine user's needs, no familiarity needed. |
| 27 | Karniouchina et al., 2008 | RBCA | Trade-off | NA | -MAD to check between predicted and actual choice<br>- RB is easier to design and estimate | 98 | No, all criteria considered at once | Multi-stage | Screen attributes choosing products, familiarity needed |
| | Asioli et al., 2013 | CBCA/DCE | Trade-off | NA | Easy to perform by DM STATA | 101 | | Multi stage | |
| | Almli et al., 2015 | Forced CBCA | Trade-off | NA | Easy to perform by DM Data analysis is harder than RBCA, but it might be software specific problem | 102 | NA | Multi stage | |

| No | Author (year) | Methods | Weight typology | Uncertainty treatment | Ease of use | Sample size | Robustness | Learning dimension | Context of use |
|---|---|---|---|---|---|---|---|---|---|
| *28* | Caputo, 2013 | CBCA/DCE | Trade off | NA | Cheap talk<br>No buy option | 200 (144 respondents are actually needed) | NA | Multi stage | Determine user's needs for food labels, no familiarity needed |
| *29* | Wezemeel et al., 2014 | CBCA/DCE | Trade-off | Confidence interval is introduced. | - Opt out alternative<br>- Cheap talk to reduce hypothetical bias<br>- SAS software support | 2400 (600 for each country) | NA | Multi-stage | Determine user's needs for nutrition label, no familiarity needed |
| *30* | Saito, 2012 | CBCA/DCE | Trade off | NA | Increase by the amount of attributes. Pre-test can be done to select attributes prior the survey | 513 | NA | NA | Determine user's motives of purchase local food, no familiarity needed |
| | Karniouchina et al., 2008 | CBCA | Trade off | - Consistency check: computation of mean absolute deviation between the predicted and actual choice. CB usually has lower validation than RB | - CB also has no-buy option that mimics real market<br>- It provides less information than a ratings task but it may be easier for respondents. | 95 | No, all criteria are considered once | Multi stage | |
| *31* | Baltussen et al., 2007 | CBCA/DCE | Multiple trade-offs among attributes | NA | Easy | 66 | NA | NA | Determine user's needs, no familiarity needed |
| *32* | Marsh et al., 2012 | CBCA/DCE | Multiple trade-offs among attributes | NA | Easy | 83 | NA | NA | |
| *33* | Defechereux et al., 2012 | CBCA/DCE | Multiple trade-offs among attributes | NA | Easy, with STATA software support | 34 | NA | NA | Screen intervention programs, familiarity needed |
| | Cleemput et al., 2014 | CBCA/DCE | Multiple trade-offs among attributes | Consistency check : good reproducibility test-retest, dominance test | - Easy if the criteria fewer than 5 or 6, amount of choices should not exceed 12.<br>- Complex data analysis | NA | No, rank reversal rarely happens where all criteria are considered at once | Multi-stage | Screen products, familiarity needed |
| | Nikou et al., 2015 | CA | Trade-off | Missing value handles with complex methods | Easy when only a few criteria are considered (< 8) | Large number of subjects required | No | Adding a new criteria | Determine user's needs, no familiarity needed |

| No | Author (year) | Methods | Weight typology | Uncertainty treatment | Ease of use | Sample size | Robustness | Learning dimension | Context of use |
|---|---|---|---|---|---|---|---|---|---|
| 34 | Elahi et al., 2016 | CF (Neighbor-based) | Trade-off | Active learning : control on the goodness of the acquired data. | -Collecting from more users<br>-cold start, i.e., when the system has not yet acquired enough ratings to generate reliable recommendations.<br>- Computer support | Large number for training data | NA | New item problem | Screen products, familiarity needed |
| | Elahi et al., 2016 | CF (Latent factor models) | Trade-off | | -Easy, based on the purchase history, computer support | Large number for training data | NA | New item problem | Screen products, familiarity needed |
| 35 | Koren & Bill, 2011 | CF (Neighbor-based) | Trade-off | Also can include the time shifting | Easy to explain the mechanism, thus also encourage users to interact with the system, fix wrong impressions, and improve long-term accuracy, computer support | Large number for training data | NA, but variability of ratings might lead to unstable data | Possible, Item-item neighbourhood models can provide updated recommendations immediately after users enter new ratings. | Screen products, familiarity needed |
| | Koren & Bill, 2011 | CF (Latent factor models) | Trade-off | inclusion of temporal dynamics to overcome the variety of time | Easy, based on the purchase history. High predictive accuracy | Large number for training data | NA | Possible, memory efficient compact model, which can be trained relatively easy | Screen products, familiarity needed |
| 37 | Desrosiers & Karypis, 2011 | CF (Neighbor-based) | Trade-off | | Item based is more useful and efficient (in terms of time, when the number of users is much greater than the number of items (amazon) | | NA, quality of data depends on ratio between the number of users and items in the system | Possible, item similarity weights could then be computed | Screen products, familiarity needed |
| 38 | Adomavicius and Tuzhilin, 2005 | CF (Neighbor-based) | Trade-off | | | | | Possible, data sparsity problem can be tackled by using user profile information when calculating user similarity. | Screen products, familiarity needed |
| 39 | Franceshet, 2011 | PageRank | Trade-off | NA | Easy, the information is obtained through purchase/ consumption history. It has computer support | | | | Screen products, familiarity needed |

| No | Author (year) | Methods | Weight typology | Uncertainty treatment | Ease of use | Sample size | Robustness | Learning dimension | Context of use |
|---|---|---|---|---|---|---|---|---|---|
| 40 | Grover, 2012 | PageRank | Trade-off | NA | Computer support | | NA | cold start | Screen products, familiarity needed |
| 41 | Diaby et al., 2016 | ELICIT | No-trade off | It is an imprecise weight elicitation | - Easy, it requires few preference information from DM<br>- Software support for PCA and Monte Carlo | 5 experts (evaluation team) | No, all criteria considered at once | Easy to re-evaluate | Screen product to finally choose one, familiarity needed |
| 42 | Danielson and Ekenberg, 2016 | Rank and surrogate weights | No-trade off | NA | Easy, the DM just has to rank. The weights are obtained from surrogate weights | NA | No, all criteria considered at once | Easy to re-evaluate | NA |