MULTIPLE CRITERIA APPROACHES FOR MEDICAL DECISION SUPPORT MODELS

MEDİKAL KARAR DESTEK MODELLERİ İÇİN ÇOK KRİTERLİ YAKLAŞIMLAR

MELODİ CEBESOY

ASST. PROF. DR. CEREN TUNCER ŞAKAR
Supervisor

ASST. PROF. DR. BARBAROS YET
Co-Supervisor

Submitted to
Graduate School of Science and Engineering of Hacettepe University
as a Partial Fulfillment to the Requirements
for the Award of the Degree of Master of Science
in Industrial Engineering.

2020

To My Family

ÖZET

MEDİKAL KARAR DESTEK MODELLERİ İÇİN ÇOK KRİTERLİ YAKLAŞIMLAR

Melodi CEBESOY

Yüksek Lisans, Endüstri Mühendisliği Bölümü
Tez Danışmanı: Dr. Öğr. Üyesi Ceren TUNCER ŞAKAR
Eş Danışman: Dr. Öğr. Üyesi Barbaros YET
Ağustos 2020, 86 sayfa

Birden fazla kriteri göz önünde bulundurarak bir dizi alternatif arasından nihai bir çözüme ulaşmak kolay değildir. Bu nedenle literatürde birçok Çok Kriterli Karar Verme (ÇKKV) yöntemi önerilmiş ve uygulanmıştır. Karar sürecine bir de belirsizlik içeren veriler dahil edildiğinde süreç daha da zorlaşır. Bu tezde, bu gibi durumlarda karar vericilere (KV) alternatiflerin değerlendirmelerini sunabilmek adına üç farklı yöntem önerilmiştir. Önerilen üç yöntem de alternatiflerin üstünlük ilişkilerine odaklanmakta ve sık kullanılan üstünlük yöntemlerinden biri olan Preference Ranking Organization Method for Enrichment Evaluation (PROMETHEE) yönteminden faydalanmaktadır. Ancak, belirsizlik içeren veriler için PROMETHEE kuralları modifiye edilmekte ve bir takım istatistiksel ve olasılıksal analizler uygulanarak karar destek için kapsamlı çıktılar elde edilmektedir. Belirsizlik içeren veriler önceki gözlemlerden, uzman değerlendirmelerinden veya Bayes Ağları (BA) olarak bilinen olasılıksal modellerden elde edilebilir. Önerilen yaklaşımlardan istatistiksel test tabanlı ve skor tabanlı olanlar sırasıyla kısmi ve tam alternatif sıralaması sunarak KV'ye farklı seviyelerde esneklik sağlamaktadır. Önerilen üçüncü yöntem olan olasılıksal PROMETHEE ise alternatiflerin her kriterdeki birleşik olasılık dağılımlarını kullanarak hem kısmi hem de tam sıralama sunmaktadır. Önerilen tüm yöntemleri test etmek için iki farklı vaka çalışması yapılmıştır. İlk olarak bir omuz ağrısı tedavi seçimi problemi üzerinde çalışılmıştır. Ardından, yöntemlerin sağlık alanı dışında kullanılabilirliğini test etmek amacıyla tedarikçi seçimi problemi uygulaması yapılmıştır.

Anahtar Kelimeler: Çok Kriterli Karar Verme, PROMETHEE, belirsizlik, Bayes Ağları, tedavi seçimi, tedarikçi seçimi

ABSTRACT

MULTIPLE CRITERIA APPROACHES FOR MEDICAL DECISION SUPPORT MODELS

Melodi CEBESOY

Master of Science, Department of Industrial Engineering
Supervisor: Asst. Prof. Dr. Ceren TUNCER ŞAKAR
Co- Supervisor: Asst. Prof. Dr. Barbaros YET
August 2020, 86 pages

It is not easy to decide on a final solution among a set of alternatives considering multiple criteria; therefore, several Multiple Criteria Decision Making (MCDM) approaches have been proposed and implemented in the literature. When uncertain data is involved in the decision process, the task gets even more difficult. In this thesis, we propose three different approaches to present evaluations of alternatives to decision makers (DMs) in such situations. All approaches focus on outranking relations of alternatives, and they utilize the well-known method Preference Ranking Organization Method for Enrichment Evaluation (PROMETHEE) in developing their evaluation measures. However, PROMETHEE rules are modified for uncertain data, and several statistical and probabilistic analyses are used to reach comprehensive outputs for decision support. The uncertain data can be obtained from previous observations, expert evaluations or samples from a probabilistic model such as Bayesian Networks (BNs). Two of the proposed approaches are the test-based and score-based approaches; they provide different levels of flexibility to the DM by offering partial and complete ranking of alternatives, respectively. The third approach, probabilistic PROMETHEE, offers both partial and complete alternative rankings using joint probability distributions of alternative evaluations in each criterion. Two different case studies are conducted to test the approaches: medical treatment selection for shoulder pain is studied to test all three approaches and also, a supplier selection case is studied to assess all approaches in a different domain. Additionally, sensitivity analyses are performed to test the sensitivity of alternative rankings to changes in criteria weights.

Keywords: Multiple Criteria Decision Making, PROMETHEE, uncertainty, Bayesian Networks, treatment selection, supplier selection

ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisors, Asst. Prof. Dr. Ceren TUNCER ŞAKAR and Asst. Prof. Dr. Barbaros YET for their unfailing support, patience, encouragements and guidance. I feel very lucky for studying with them throughout the project.

I am grateful to my family for their endless love and support in every moment of my life. I also thank all my close friends who are with me whenever I need.

I would like to thank TUBITAK for supporting my academic work. This work was supported by an Institutional Links grant, ID 352394702, under the Turkey - Newton - Katip Çelebi Fund partnership. The grant is funded by the UK Department of Business, Energy and Industrial Strategy (BEIS) and The Scientific and Technological Research Council of Turkey (TUBITAK - ID 217M518) and delivered by the British Council. For further information, please visit www.newtonfund.ac.uk.

Melodi CEBESOY
August 2020, Ankara

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

NOTATIONS AND ABBREVIATIONS

Notations

a_i	Solution *i*
d_{ikj}	Difference between evaluations of a_i and a_k on criterion *j*
w_j	Weight of criterion *j*
$P_j(d_{ikj})$	Preference value of a_i on a_k
c_{ik}	Aggregated preference index of a_i over a_k
φ_i^+	Positive flow of a_i
φ_i^-	Negative flow of a_i
φ_i	Net flow of a_i
φ_{is}^+	Positive flow of a_i in sample *s*
φ_{is}^-	Negative flow of a_i in sample *s*
$\overline{\varphi_i^+}$	Mean of positive flows of a_i
$\overline{\varphi_i^-}$	Mean of negative flows of a_i
S	Set of all samples
S_{it}	Subset of S where a_i occupies the t^{th} rank
φ_{is}	Net flow of a_i in sample *s*
r_{is}	Rank of a_i in sample *s*
w_t	Weight of rank *t*
p_{it}	Probability of a_i having the t^{th} rank
$F_{r_i}(x)$	Cumulative distribution of a_i for all possible ranks
θ_i	Weighted score of a_i
r_i^{score}	Rank of a_i provided by the score-based approach
J_{ik}^j	Joint probability distribution of a_i and a_k in criterion *j*
P_{ik}^j	Dominance score of a_i on a_k in criterion *j*
M_{ik}	Weighted dominance score of a_i on a_k
WP_{ik}^j	Weighted probabilistic score of a_i on a_k in criterion *j*
s_j	Number of levels of criterion *j*
d	Criterion level distance between a_i and a_k

Abbreviations

AHP	Analytic Hierarchy Process
BN	Bayesian Network
DM	Decision Maker
ELECTRE	Elimination and Choice Expressing the Reality
MCDM	Multiple Criteria Decision Making
MCS	Monte Carlo Simulation
PROMETHEE	Preference Ranking Organization Method for Enrichment Evaluation
ROC	Rank Order Centroid
RR	Rank Reciprocal
RS	Rank Sum

1. INTRODUCTION

Multiple Criteria Decision Making (MCDM) is a domain of methods that model and solve problems where multiple criteria are simultaneously considered. A criterion is a performance measure that needs to be defined and measured to analyze and compare decision alternatives. Since real-life problems usually do not have a single objective, every day we must often decide considering different criteria. However, generally it is not possible to optimize all criteria at the same time. For example, when buying car, we consider some criteria like comfort level, cost, safety and color. It is very difficult to find a car which fits a low budget and has high comfort and safety standards at the same time. In such cases, MCDM methods are used to assist decision makers (DMs).

If MCDM problems involve uncertainty, the situation becomes more complicated. The uncertainty can be caused by internal or external uncertainty. Internal uncertainty occurs with ambiguity due to model structure or inputs. On the other hand, external uncertainty is caused by lack of knowledge on solution results (Stewart, 2005).

In MCDM terminology, a solution x is called efficient if it is impossible to improve a criterion without worsening at least another one. Inefficient solutions should not be considered by DMs, however three general MCDM problem types have emerged for assessing efficient solutions: choice, sorting and ranking (Roy, 2005). Approaches in choice problems aim to select the best alternative within the acceptable solutions; sorting approaches assign solutions into preference-ordered categories, and ranking approaches aim to list the solutions from the best to the worst in complete or partial rankings. In this thesis, outranking relations of alternatives with uncertain criteria evaluations are studied which are used to choose from, sort or rank alternatives. Outranking relations are evaluated by comparing alternative solutions in pairs for each criterion and aggregating those comparisons with preference measures. One of the most widely known outranking approaches is The Preference Ranking Organization Method for Enrichment Evaluation (PROMETHEE), and this thesis uses the outranking structure of PROMETHEE and extends it to work under uncertainty. PROMETHEE is able to provide different types of preference functions for comparisons, handle different types of criteria and create both complete and partial rankings of alternatives.

We propose different approaches to find alternative rankings for MCDM problems where criteria evaluations of solutions involve uncertain data. This data that includes uncertain criteria evaluations may be gathered by a data-generating model, expert evaluations or previous observations. We propose three methodologies to handle such cases. The first two methodologies use sampling from probability distributions to cope with uncertainty and modify PROMETHEE to work with samples; these are the test-based and score based outranking approaches. The test-based approach provides partial ranking whereas the score-based approach provides full ranking of solutions. The test-based approach enables the DM to examine the error rates of MCDM analysis, and the score-based approach provides a score calculated from the probabilities of solution rankings and a graphical tool to assess their uncertainty. On the other hand, the third methodology handles uncertainty using joint probabilities of solution pairs instead of sampling and uses modified PROMETHEE scores to provide partial and full rankings. We provide three variations of this third methodology for different types of DM preferences. We apply our methodologies on healthcare problems involving uncertainty regarding the outcome of treatment options for specific patients with known conditions. We specifically use healthcare examples since this thesis was conducted as a part of the project "Interdisciplinary Research Links for Medical AI: Management of Musculo-skeletal Injury" which is supported under the Turkey – Newton - Katip Çelebi Fund partnership. We also include some applications on the multiple criteria supplier selection problem to demonstrate the methodologies in a different area.

In this thesis, we propose using Bayesian Networks (BN) models to represent uncertain data. BN models provide probability distributions of alternative solutions for each criterion. Using BN models with MCDM is suitable since they provide flexible representations of uncertainty. This thesis can also be useful in presenting a guideline for the use of BN and MCDM approaches together.

In Section 2 of the thesis, we review the literature on handling uncertainty in MCDM and MCDM in healthcare. In Section 3, necessary background information on PROMETHEE, preference elicitation and BNs are provided. Section 4 includes our proposed methodologies. In Section 5, we apply our approaches to different problems and present computational results. Section 6 provides conclusions and discussions.

2. LITERATURE REVIEW

This section of the thesis reviews the MCDM methods that have been used for problems with uncertainty, and MCDM approaches applied in healthcare.

2.1. MCDM Approaches under Uncertainty

There is a vast literature on MCDM methods, but since we focus on problems involving uncertainty, our review also focuses on such problems. For more comprehensive reviews of MCDM problems with uncertainty, the readers are referred to the following papers: Durbach and Stewart (2012) reviewed the literature on MCDM under uncertainty based on five uncertainty representations which are probabilities, decision weights, risk measures, fuzzy numbers and scenarios. Probability based uncertainty was handled using multi-attribute utility theory (MAUT), pairwise comparisons of distributions, belief functions or simulating from probability distributions. Outranking relations and stochastic dominance are the most common approaches which use pairwise comparisons. Another widely used method is Monte Carlo Simulation (MCS), it is reviewed under the models simulating from distributions. Additionally, decision problems with fuzzy numbers are generally modelled using weighted additive sums, Analytic Hierarchy Process (AHP), comparisons to ideal solutions and fuzzy or rough sets. Broekhuizen et al. (2015) reviewed approaches for Multi Criteria Decision Analysis (MCDA) under uncertainty in healthcare. They analyzed five uncertainty approaches: Bayesian framework, deterministic sensitivity analysis, probabilistic sensitivity analysis, fuzzy set theory and grey theory for different MCDA methods. They concluded that in most of the healthcare decision problems, deterministic sensitivity analysis is more advantageous due to its easy implementation and it is widely studied with AHP in the literature.

To incorporate uncertainty in MCDM problems, MCS is one of the well-known approaches. Many previous papers used MCS for sampling from probability distributions. Baudry, Macharis and Vallee (2018) employed MCS within Multi Actor Multiple Criteria Analysis (MAMCA). They applied AHP to derive criteria weights in each stakeholder group. Stakeholder groups were equally weighted. They performed a case study to select the best biofuel option. After determining probability distributions of alternative solutions in each criterion, by using MCS, their new solution framework provided probabilities of each rank instead of providing a single ranking. Momani and Ahmed (2011) also performed a hybrid model that uses AHP and MCS together. They aimed to select the best material handling equipment when there is an uncertainty caused by human preferences. Their proposed model differs from the traditional AHP procedure by making pairwise comparisons with random variables. Using MCS, 1000 replications were generated for each pairwise comparison. Then, results were used to find alternative weights with AHP. They implemented the proposed model in a pharmaceutical plant. Betrie et al. (2013) presented a study that uses MCS and PROMETHEE to make decisions in the face of uncertainty. They proposed two approaches: deterministic and probabilistic analysis. In the deterministic case, they used basic PROMETHEE rules and AHP to derive criteria weights. In the probabilistic case, they generated criteria weights using AHP and then, they fitted a probability distribution. To handle the uncertainty that comes from the random variables of weights, they used MCS. Also, they made sensitivity analysis to see the effect of each criterion on the final ranking. To do that, they utilized Spearman rank correlation coefficient. Their methodology was performed on a case of selecting mine sites. The purpose of Levary and Wan (1998) was to deal with two types of uncertainty when deciding in a multiple criteria problem using AHP. The first type of uncertainty came from lack of knowledge to compare some criteria pairs. The other type of uncertainty was caused by ambiguous future events of the decision problem. To illustrate the proposed methodology, they used a university selection case for a Ph.D. graduate. They used a range by determining distributions for some pairwise comparisons of criteria and performed scenario analysis to handle the uncertainty that comes from future events. Hopfe, Augenbroe and Hensen (2013) modified the traditional AHP procedure to compare two building designs considering probabilistic performance values. They conducted building performance simulation for 200 replications. They obtained ranges of performance evaluations for criteria. Then, a new adapted AHP score was calculated to determine the best alternative. Dorini, Kapelan and Azapagic (2011) focused on three cases: no uncertainty exists, model or data has uncertainty and preferences have uncertainty. They applied compromise programming and MCS on a case study about comparing two sustainable electricity generation options. For the cases which contain uncertainty, they fitted probability distributions and performed MCS. From the results, they reached probabilities of an alternative being better than the other. Therefore, they presented the better option found with a confidence rate. Baležentis and Streimikiene (2017) used additive ratio assessment, weighted aggregated sum and TOPSIS to evaluate effective energy planning scenarios. They had conflicting criteria under technical, environmental and economic concepts. After solving the problem with three approaches, they made sensitivity analysis for rankings using different criteria weights using MCS.

The other widely used approach for handling uncertainty in MCDM problems is fuzzy theory. Montazar, Gheidari and Snyder (2013) proposed a hybrid approach that uses both fuzzy triangular numbers and AHP. They aimed to make performance assessments of irrigation alternatives with uncertain parameters. Twenty-one sub criteria were used to evaluate four alternatives under environmental, technical, social and economical concepts. To generate weights of criteria, fuzzy AHP was used. Benetto, Dujet and Rousseaux (2008) proposed a fuzzy multiple criteria approach to modify an existing multi criteria method called NAIADE with the aim of improving life cycle assessments. They evaluated alternatives by fuzzy pairwise comparisons with uncertain evaluations. Kilic, Zaim and Delen (2014) performed a three-phase decision-making method to select an appropriate ERP system for Turkish Airlines. Because the criteria priorities varied according to the people in the process, they applied fuzzy theory and AHP. Then, they employed TOPSIS to rank the alternatives. Montazer, Saremi and Ramezani (2009) focused on two important steps of decision making: performance evaluations of alternatives and alternative rankings. They presented a fuzzy evaluation approach to assess alternative performances and to rank them, they integrated fuzzy theory to ELECTRE III to construct outranking relations in an uncertain environment. They implemented their approach to a supplier selection problem in oil industry. Pitchipoo, Venkumar and Rajakarunakaran (2013) introduced an integrated model of fuzzy AHP and grey relational analysis (GRA) for a selection problem of suppliers. Uncertainty about criteria weights was modeled with fuzzy set theory and the alternatives were ranked with the help of GRA. Additionally, they conducted sensitivity analysis regarding grey equation coefficients. They obtained that ranking of supplier alternatives stay stable by changing the coefficients. Therefore, they claimed that they reached a robust solution. Kuang, Kilgour and Hipel (2012) used grey theory to deal with uncertainty on criteria weights and different DMs. PROMETHEE II was used to produce a ranking of alternative source water protection strategies. They made the analysis with six criteria and eight alternatives.

Another group of approaches to assess alternatives under multiple criteria when uncertainty exists is the stohactic dominance concept. Zhang, Fan and Liu (2010) and Liu, Fan and Zhang (2011) used the stochastic dominance concept in developing methodologies to rank alternatives with uncertain criteria outcomes. They created dominance matrices based on stochastic dominance degrees, and then they used these matrices with PROMETHEE to find the rank order of alternatives. A similar study was performed by Rietveld and Ouwersloot (1992) to rank location alternatives of nuclear plants. Their methodology is appropriate for both ordinal and mixed data to conduct stochastic dominance rules.

Some other approaches considering uncertainty in MCDM problems involve Hyde, Maier and Colby (2005) who introduced a distance-based approach to handle the uncertainty in criteria weights. Their method presents a new approach to sensitivity analysis to overcome the limitations of existing sensitivity analysis techniques for weights. D'Avignon and Vincke (1988) presented a distributive multiple criteria approach for alternatives that have probabilistic evaluations on each criterion. They made ranking decisions based on distributive strengths and weaknesses of alternatives. Durbach (2014) aimed to extend the distributive approach of D'Avignon and Vincke (1988) by simulating scenarios.

In our proposed approaches, we use PROMETHEE, therefore we also provide a focused review of handling uncertainty in outranking methods. Early work in this area focuses on modeling the uncertainty and variation between multiple experts. Mareschal (1986) proposed a methodology using PROMETHEE I and II with the average ranking of different experts to reach the final decision in a project selection problem. Different statistical and operational research techniques have also been used to model uncertainty in different elements of PROMETHEE. Another study was conducted by Beynon and Wells (2008) using PROMETHEE II and an evolutionary algorithm to analyze the minimum amount of change required in criteria values to improve the preference rank of alternatives. Yuen and Ting (2012) used a fuzzy approach with PROMETHEE II to cope with uncertainty in decision alternatives. They applied their approach on a textbook selection case. Cavalcante and De Almeida (2007) focused on expected values of criteria by using a Bayesian approach, and applied PROMETHEE III on those values to decide about maintenance planning alternatives.

MCS has been a widely used approach to include uncertainty in PROMETHEE calculations too. Hyde, Maier and Colby (2003) fitted probability distributions to model criteria weights. PROMETHEE II scores were used to rank the alternatives. They used MCS to run the models and reported the probability of alternatives occupying different ranks. They performed a case study to select between six renewable energy alternatives. Doumpos and Zopounidis (2010) made sensitivity analysis of PROMETHEE II parameters and criteria preferences using MCS. Gervásio and Simões Da Silva (2012) fitted probability distributions to criteria and used MCS and PROMETHEE II to calculate the probabilities of alternatives holding each possible rank. However, they did not propose a measure to aggregate these probabilities and to obtain a final decision.

2.2. MCDM Approaches in Healthcare

Since we mainly apply our proposed approaches to healthcare problems, we also review papers using MCDM approaches in decision making problems in medical area. Many researchers studied the problem of selecting the best alternatives. Li et al. (2018) proposed a methodology based on heterogeneous MCGDM in order to decide on the best alternative when the information of doctor and patients do not match. Özkan (2013) investigated the status of clinical waste management and selected the most appropriate out of five disposal methods with the use of ANP and ELECTRE. Kulak, Goren, and Supciller (2015) developed RFAD method, a new MCDM approach which includes risk factors in the decision process and they used it for selection of medical imaging devices.

Some researchers studied prioritization of different subjects in healthcare. Zhang et al. (2018) developed an intuitionistic multiplicative ORESTE method to assign order of patients for hospitalization based on complex and related factors. Nobre, Trotta and Gomes (1999) evaluated alternatives based on criteria using TODIM (Tomada de Decisao Interativa Multicriteo). Their aim was to support decision process for patients when their behavior is uncertain. Taghipour, Banjevic, and Jardine (2011) studied on the prioritization of maintenance of medical equipment. They applied AHP to develop a maintenance strategy based on some criteria such as function, age, risk, task criticality, maintenance requirements and hazard warning.

Another research was made on decision analysis in oncology. Adunlin et al. (2015) determined harm-benefit balance of cancer treatments. They used ELECTRE for decision making based on evidences under complex situations.

Some papers were based on analysis and comparison of different MCDM approaches. Sałabun and Piegat (2017) compared COMET, TOPSIS and AHP based on the evaluations of mortality of patients with acute coronary syndrome. They concluded that COMET method gives more accurate solutions in opposition to TOPSIS and AHP. Additionally, Dolan and Veazie (2018) investigated whether increase of multi criteria decision support is able to provide more efficient decisions for the community and it leads to a decrease in the ease of use. It is found that increase of multi criteria decision support does not cause a decrease in ease of use, MCDM based systems can be developed for clinical treatments.

Souissi et al. (2017) used ELECTRE and MR-Sort to prescribe antibiotics according to other medical conditions of patients such as allergy and kidney diseases. In conclusion, they categorized antibiotics into 3 groups based on recommendation rates and developed a system for doctors to be efficient in prescribing. Dong et al. (2014) created clusters of alternatives for pulsation data and used TOPSIS to select the proper ones within them. They concluded that studying pulsations in 13 clusters is the most suitable one. Rahimi, Gandy, and Mogharreban (2007) developed a web based medical diagnosis system that is able to provide information-based updates, short response time and ease of use.

The purpose of some researches was to analyze the performances of treatments or services. Kuo, Wu, and Hsu (2012) aimed to improve the service of elderly outpatients. With TOPSIS and fuzzy clustering theory, they ranked failure risks in Healthcare Failure Mode Effects Analysis (HFMEA). Lupo (2016) presented an AHP based study to procure reliable estimation of service quality. Their proposed method was applied to the service quality of Sicilian hospitals and causes of patient dissatisfaction were found. Another MCDM model was developed by Ghosh (2008) to rank the factors that affect the surgical performance of the hospital by Fuzzy Composition Programming (FCP). Nilashi et al. (2016) used Fuzzy ANP in order to find the most important factors among technology, environment, human and organization for the hospital information system adoption of hospitals in Malaysia. Also, La Scalia et al. (2011) investigated the probability of success of transplantation of pancreas islet. Applying fuzzy TOPSIS, three samples are evaluated considering four factor groups: donor, pancreas, islent and recipient.

Diaby and Goeree (2014) reviewed the use of MCDA methods in medical decision making. They categorized methods into five groups: elementary (Maximin, Hurwicz), value-based (multi-attribute value theory (MAVT), MAUT, AHP), goal programming, reference methods and outranking methods (ELECTRE versions, PROMETHEE). Additionally, they implemented some of the methods with hypothetical cases.

In this thesis, we introduce three different approaches to provide decision support for multiple criteria problems with uncertain data. In the literature, there are several applications where uncertainty is handled in MCDM problems; however, our approaches differ from them. Our approaches can work with uncertain data generated by different sources such as past observations, samples from probabilistic models or judgements of multiple experts. Posterior probabilities derived by a BN model can be used for alternative evaluations in criteria. Thanks to this approach, we also offer general approaches for the use of MCDM and BN posteriors. With our test-based approach, we consider error rates caused by the uncertainty and present partial ranking of alternatives. This approach is advantageous for situations where the DM desires some level of flexibility. With our score-based approach, using probabilities of alternatives occupying for each rank, we present a final solution for both risk-averse and risk-seeking DMs. Our third approach uses probabilities directly instead of creating samples from them. Thus, it provides a novel modified probabilistic approach to the classical PROMETHEE. Furthermore, in medical area, our approaches propose patient-specific decision support using BN model posteriors.

3. BACKGROUND INFORMATION

We propose MCDM approaches based on outranking relations of alternative solutions in the face of uncertain criteria values. Our approaches use PROMETHEE scores as underlying performance evaluations, but first we need to determine weights of criteria to express DM preferences and also mechanisms to make PROMETHEE work with uncertain criteria outputs. Before explaining our approaches, in this section we give background information on MCDM, outranking relations and rules of PROMETHEE, weight elicitation techniques used to determine and express the importance of criteria, and BNs that provide probability distributions for criteria values.

3.1. PROMETHEE

The aim of MCDM is to consider multiple criteria simultaneously in assessing and deciding about alternative solutions. MCDM methods eliminate inefficient solutions from consideration and focus on studying efficient solutions. In MCDM terminology, a solution *x* is called efficient if there is no other solution *y* that is better than or equal to *x* in all criteria, and strictly better than *x* in at least one criterion. Within the general scope of MCDM, outranking methods compare alternative solutions in pairs and reach final performance measures that are used to determine partial or full ranking or sorting of solutions.

PROMETHEE is an outranking approach proposed by Brans, Vincke and Mareschal (1986), and it is one of the most widely-used outranking methods. Basically, it makes pairwise comparisons of solutions in terms of each criterion in the decision problem. Based on different preference functions, it assigns a preference value between 0 and 1 to each solution pair in comparison. Using these preference values and criteria weights, an aggregated measure is calculated to determine preferability degree of solutions. Let us assume that a_i and a_k are two solution alternatives to be evaluated under *m* maximization type (without loss of generality) criteria and their evaluations are $a_i = (a_{i1}, a_{i2}, \dots, a_{im})$ and $a_k = (a_{k1}, a_{k2}, \dots, a_{km})$. To determine the preference strength of a_i over a_k, first of all, the magnitude of difference between evaluations of a_i and a_k is calculated for each criterion *j* by $d_{ikj} = a_{ij} - a_{kj}$. After that, depending on the magnitude of this difference, a preference value is determined using one of the preference functions given in Figure 3.1. The function type is selected by the DM for each criterion. For simplicity, difference between evaluations is represented by d_{ik} since every criterion has a similar procedure. Some preference functions have threshold values q and p. The indifference threshold q represents the minimum required difference for the DM to express any level of preference, and the preference threshold p represents the minimum required difference for the DM to express strict preference. These thresholds are specific to each criterion and DM.

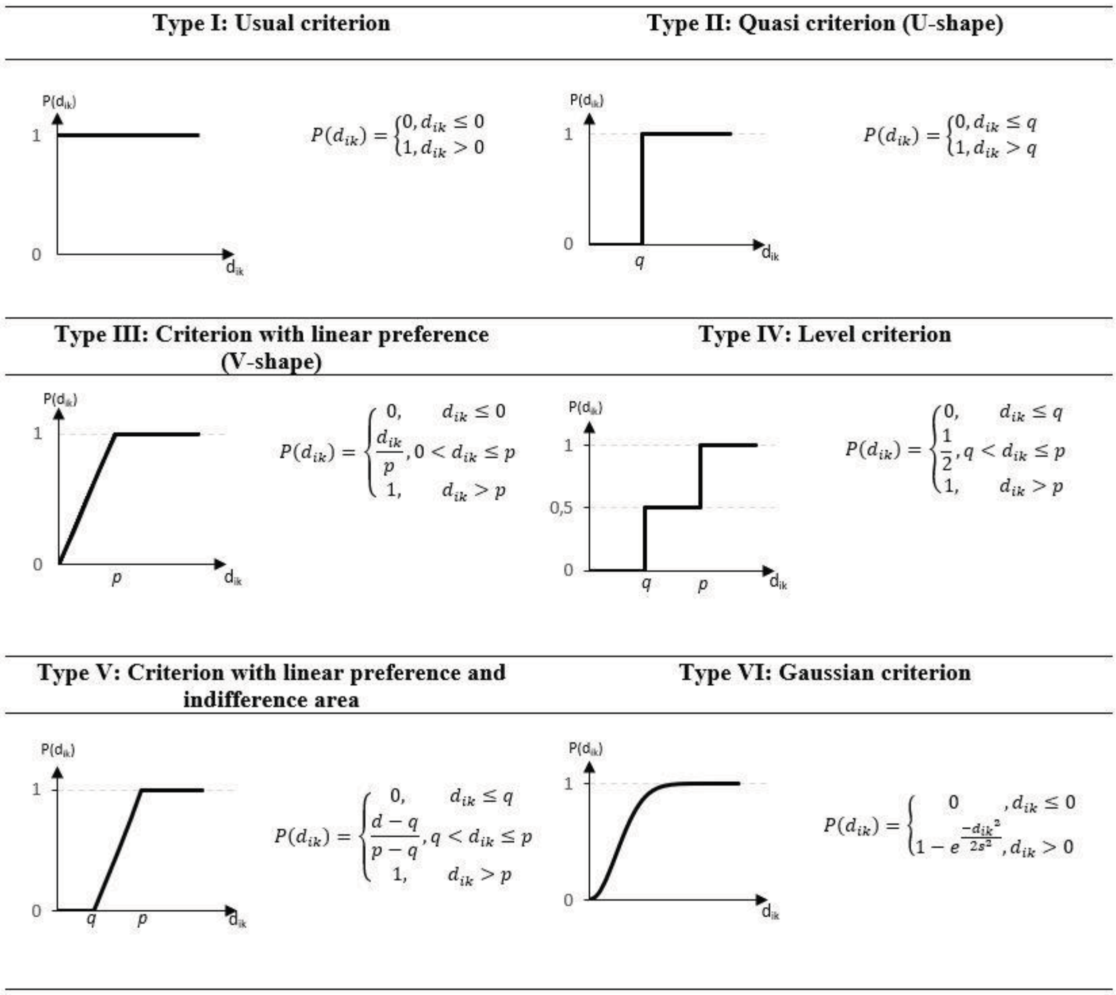


Figure 3.1. Preference functions

The preference value of a_i over a_k on criterion j is represented by $P_j(d_{ikj})$ and w_j denotes the weight of criterion j. The weights in PROMETHEE correspond to the importance levels of the criteria for the DM. To find overall preference value of a_i over a_k regarding all criteria, aggregated preference index of a_i over a_k is calculated using (1).

$$c_{ik} = \sum_{j=1}^{m} P_j(d_{ikj}) w_j \tag{1}$$

Criteria weights used in PROMETHEE must be pre-specified, weight elicitation techniques for this task that we review are provided in Section 3.2. Also, weights should sum up to 1. Therefore, c_{ik} values are between 0 and 1. Values closer to 0 represent weak global preference of a_i over a_k and values closer to 1 represent strong global preference of a_i over a_k.

After all pairwise comparisons are completed, PROMETHEE calculates overall preference indices for all solutions. Positive flow and negative flow of a_i are denoted by φ_i^+ and φ_i^- respectively. Positive flow implies how strongly a_i outranks all other solutions and it is calculated as in (2). On the other hand, negative flow of a_i implies how strongly other solutions outrank a_i and it is calculated using (3). A solution with higher positive flow and lower negative flow is considered to be preferable.

$$\varphi_i^+ = \sum_k \mathrm{c_{ik}} \tag{2}$$

$$\varphi_i^- = \sum_k \mathrm{c_{ki}} \tag{3}$$

There are a few versions of PROMETHEE but two of them dominated the literature: PROMETHEE I and II. They work with the positive and negative flows that we defined. PROMETHEE I evaluates solutions considering positive and negative flows simultaneously.

In PROMETHEE I, a_i is preferred to a_k if one of the following conditions hold:

(i) $\varphi_i^+ > \varphi_k^+$ and $\varphi_i^- < \varphi_k^-$

(ii) $\varphi_i^+ = \varphi_k^+$ and $\varphi_i^- < \varphi_k^-$

(iii) $\varphi_i^+ > \varphi_k^+$ and $\varphi_i^- = \varphi_k^-$

Solutions a_i and a_k are incomparable if one of the following conditions hold:

(i) $\varphi_i^+ > \varphi_k^+$ and $\varphi_i^- > \varphi_k^-$

(ii) $\varphi_i^+ < \varphi_k^+$ and $\varphi_i^- < \varphi_k^-$

Lastly, there is an indifference relation between a_i and a_k if $\varphi_i^+ = \varphi_k^+$ and $\varphi_i^- = \varphi_k^-$.

As a result of PROMETHEE I, it may not be possible to achieve full ranking of solutions since there can be incomparability or indifference between some pairs of solutions.

On the other hand, PROMETHEE II works with a final score as calculated in (4) and it is called the net flow of solution a_i. It is used to achieve a full ranking of solutions.

$$\varphi_i = \varphi_i^+ - \varphi_i^- \tag{4}$$

Solutions are ranked in decreasing order of their net flows. Therefore, PROMETHEE II allows a strict rank list, however it is possible to lose some information that comes from positive and negative flows by calculating (4). The following steps summarize the PROMETHEE approach.

1. For each criterion j, difference between evaluations of a_i and a_k is calculated by $d_{ikj} = a_{ij} - a_{kj}$.
2. For each criterion j, DM selects a preference function from Figure 3.1 and the related parameters, if any.
3. For each alternative pair (a_i, a_k) in each criterion j, a preference value $P_j(d_{ikj})$ is calculated.
4. For each alternative pair (a_i, a_k), aggregated preference index c_{ik} is calculated using (1).
5. For each a_i, positive and negative flows are calculated using (2) and (3).
6. For each a_i, net flow is calculated using (4).

7. Obtain an outranking relation of alternatives:
   7.1. To obtain partial ranking of alternatives, PROMETHEE I rules given above are used.
   7.2. To obtain complete ranking of alternatives, list all alternatives in decreasing order of net flows.

PROMETHEE is a flexible and powerful methodology to compare and rank solutions. It can work with continuous and discrete measurements of criteria, as well as ordinal and binary. The DM can express preferences through different preference functions and can fine-tune those preferences with customized indifference and preference thresholds. Moreover, these thresholds are not abstract values like the concordance and discordance thresholds used by the other popular outranking method, ELECTRE. In PROMETHEE, the DM can realistically be expected to express the minimum difference in the values of a criterion so that there will be any significant difference, or absolute superiority. Due to these advantages, we work with PROMETHEE in our proposed approaches.

3.2. Weight Elicitation Techniques

Since PROMETHEE, like many other MCDM methods, does not have a weight derivation step in its procedure, other techniques should be used to elicit weights before executing PROMETHEE. Therefore, in this section, we will review some of these techniques used in the literature. Methods are categorized as the ones based on criteria comparisons and based on alternative comparisons.

3.2.1. Methods Based on Comparison of Criteria

3.2.1.1. Direct Evaluation Methods

Direct evaluation methods are the most widely known weight elicitation methods because they are the oldest ones. Detailed explanations and examples are provided in Pomerol and Barba-Romero (2000). We explain three main methods in this category.

Simple Ranking

Simple ranking is the easiest way of deriving criteria weights. The DM is just asked to rank criteria considering his/her preferences for the problem. Then, scores are assigned to criteria in accordance with the ranking. For example, the criterion which is the least important takes score 1, the next one takes score 2 and so on. The criteria with equal importance take the average of scores like 3.5 instead of 3 and 4. After that, all criterion scores are divided by the sum of scores and criteria weights are obtained.

The method is advantageous due to its simplicity and low requirements for calculations, however it does not allow the criteria to have all possible values between 0 and 1 and its effectiveness is not proven.

Simple Cardinal Evaluation

Simple cardinal evaluation is another simple way of weight elicitation. The difference between this and the previous method is that DM is asked to assign scores to the importance of criteria on given scales like 0-5 or 0-100. Scoring is done by considering that the most important one should take the highest score. After that, each criterion score is divided by the sum of scores to elicit weights. For example, let us assume that we have three criteria with scores 5, 3 and 2. Then, their weights will be 0.5, 0.3 and 0.2 since the sum of scores is 10.

This method asks DM for more information than simple ranking, but the results heavily depend on the scale used and the DMs tend to change their answers with repetitions.

The Method of Successive Comparisons

The first version of this approach goes a long way back and other versions were proposed within years (Pomerol and Barba-Romero, 2000). This method requires more information from the DM, but it gives more consistent results than the simple cardinal evaluation method. Basically, the procedure can be applied in the following steps.

Firstly, criteria are ranked by DM and they are assigned a score using a scale as in the previous cardinal evaluation method. After that, starting from the first one, criteria are compared with their consecutive ones. For example, the first criterion is compared with the second criterion then, the first criterion is compared with the second plus third criterion, and so on. These successive comparisons are made for each criterion. In the next step, the consistency between the comparisons and the scores are checked.

Inconsistent scores based on comparisons are revised. Then scores are normalized, and weights are obtained.

3.2.1.2. Ranking-based Methods

Ranking based methods are another group of procedures for deriving weights. Criteria rankings provided by DM are used to calculate weights; but the procedures are somewhat more sophisticated than simple ranking. The most common formulas are the rank sum (RS), rank reciprocal (RR) and rank order centroid (ROC).

RS approach assumes equal distance between weights of consecutive ranks while RR and ROC approaches increase the difference between consecutive weights as the rank positions get higher. Formulas are given in (5), (6) and (7) where $j=1,2,\ldots,N$ and r_j is the rank of criterion j.

$$w_j(RS) = \frac{N - r_j + 1}{\sum_{k=1}^{N} N - r_k + 1} \tag{5}$$

$$w_j(RR) = \frac{1/r_j}{\sum_{k=1}^{N}(1/r_k)} \tag{6}$$

$$w_j(ROC) = \frac{1}{N}\sum_{k=j}^{N}\frac{1}{r_k} \tag{7}$$

Among these formulas, ROC has the best performance with respect to choice accuracy. For more detailed discussions, Sureeyatanapas (2016), Roszkowska (2013) and Ahn (2011) can be reviewed.

3.2.1.3. ELICIT

Diaby, Sanogo and Moussa (2016) proposed ELICIT as a new approach for weight elicitation for multiple DMs and applied it to a healthcare problem. ELICIT has two fundamental steps. In the first step, each DM assigns a rank to each criterion based on their individual preferences. To aggregate these rankings, a statistical approach is used. Firstly, for each ranking provided by m different DMs, (8) is used to calculate standardized ranks of n criteria and a new standardized matrix (S_{mxn}) is constructed regarding standardized values. R_s and R_p represent the standardized and provided ranks of each criterion. Besides, μ and σ are mean and standard deviation of the rankings of each DM.

$$R_s = \frac{R_p - \mu}{\sigma} \tag{8}$$

In the next step, correlation matrix is created by (9).

$$CM = \frac{SxS^T}{n} \tag{9}$$

Using the correlation matrix, eigen value, corresponding eigen vector and normalized eigen vector (μ_{norm}) are calculated, then they are used to find criteria scores as in (10).

$$Score = S^T \mu_{norm} \tag{10}$$

Criteria weights are listed in decreasing order of scores. To elicit criteria weights from that order, MCS is used. 1000 iterations are simulated under two conditions; the pre-defined order of weights should hold and summation of all criteria weights should be equal to 1. As a result of that simulation, criteria weights are determined.

3.2.1.4. Stepwise Weight Assessment Ratio Analysis (SWARA)

The method SWARA is proposed by Keršuliene, Zavadskas and Turskis (2010) to derive weights considering DM's opinions about significance ratio of criteria. First, the DM is asked to provide a ranking of criteria regarding the available information so that the first is the most important one for the case. Then, each criterion *j* is compared with the criterion *j+1* by the DM to find how much criterion *j* is relatively more important than criterion *j+1* and this relative importance value for each *j* is defined as s_j. After each relative importance between criteria is found, a coefficient value (k_j) for each criterion *j* is determined using (11). Weights of each criterion is found using (12) and normalized using (13).

$$k_j = \begin{cases} 1, & if\ j = 1 \\ s_j + 1, & if\ j > 1 \end{cases} \tag{11}$$

$$w_j = \begin{cases} 1, & if\ j = 1 \\ \dfrac{w_{j-1}}{k_j}, & if\ j > 1 \end{cases} \tag{12}$$

$$w_j^{norm} = \frac{w_j}{\sum w_j} \tag{13}$$

3.2.1.5. The Analytic Hierarchy Process (AHP)

The AHP is an MCDM approach proposed by Saaty (1990). It is widely used for both weight elicitation and alternative ranking in multiple criteria decision problems. Since we use PROMETHEE to rank alternatives, we focus on the criteria weighting procedure of AHP. Generally, the process is based on pairwise comparisons according to an importance scale. Therefore, it makes it possible to express qualitative judgements as numeric values.

In the first step, alternatives and criteria of the problem are defined hierarchically. Thus, the aim of the problem, criteria which affect the solution and solution alternatives are shown clearly for the comparison process.

At the second step, all criteria are compared in a pairwise manner. The DM determines which criterion in the pair is more important and the numeric equivalent of that importance using the scale in Table 3.1.

Table 3.1. Importance scale for criteria comparisons in AHP

1	Equally preferred
3	Moderately preferred
5	Strongly preferred
7	Very strongly preferred
9	Extremely preferred
2,4,6,8	Intermediate values

In the presence of m criteria, let the DM compare criteria j and l and choose j as the more important one. If the score of this importance is denoted by h_{jl}, then h_{lj} directly becomes

$1/h_{jl}$. Using this rule, all comparisons are completed and for all criteria pairs a normalized comparison score is calculated using (14).

$$\bar{h}_{jl} = \frac{h_{jl}}{\sum_{n=1}^{m} h_{nl}} \tag{14}$$

To derive criteria weights, normalized comparison scores are averaged using (15).

$$w_j = \frac{\sum_{n=1}^{m} \bar{h}_{jn}}{m} \tag{15}$$

With these steps, criteria weights which sum up to 1 are obtained and a consistency test is conducted using the steps below. Since several pairwise comparisons are performed in AHP, it is important that the DM makes these comparisons consistently with each other. For example, if the first criterion is 2 points more important than the second criterion and the second criterion is 3 points more important than the third criterion, then the first criterion should be around 6 points more important than the third criterion. The critical threshold for consistency is accepted as 0.10. If consistency ratio is higher than this threshold, then the DM should review the importance levels assigned and repeat the process. Steps of calculating the consistency ratio are given in the following steps:

1. For the comparison matrix consisting of h_{jl} and h_{lj} values, multiply each value in each column with the weight of the corresponding criterion, then sum these vectors to obtain a weighted sum vector.
2. Divide each value in the weighted sum vector by the weight of the corresponding criterion.
3. Take the average of values found in the previous step and define it as λ.
4. Calculate the consistency index (CI) using (16), where m is the number of criteria compared.

$$CI = \frac{\lambda - m}{m - 1} \tag{16}$$

5. Calculate the consistency ratio (CR) using (17). RI is the consistency index of a randomly generated pairwise comparison matrix, it is based on the number of criteria compared and given in Table 3.2.

$$CR = \frac{CI}{RI} \tag{17}$$

Table 3.2. RI values for different number of criteria

m	3	4	5	6	7	8
RI	0.58	0.90	1.12	1.24	1.32	1.41

3.2.2. Methods Based on Comparison of Alternatives

3.2.2.1. The UTA Methods

The UTA (Utility Additive) Method was presented by Jacquet-Lagreze and Siskos (1982) to evaluate additive utility functions of alternatives considering multiple criteria. It requires the DM to rank some alternatives subjectively to obtain preference information. Then, it uses linear programming to obtain utility function parameters that obey the preferences of the DM, which in turn are used to provide rankings of all alternatives. UTA is not basically proposed to elicit weights, but it provides some estimations of weights as a by-product. There are some versions of the UTA method in the literature such as UTASTAR and UTADIS. UTASTAR was proposed by Siskos and Yannacopoulos (1985) to enhance the UTA method to improve its performance. Another method developed using UTA is the UTADIS method which is introduced by Jacquet-Lagreze, (1995). Differently from the UTA, UTADIS requires the DM to rank the alternatives in groups. So, the best alternatives are assigned to the first group.

3.2.2.2. A Mathematical Modelling-based Method

The method proposed by Tuncer Şakar and Yet (2018) is based on the alternative ranking of the DM in terms of his/her preferences. It does not require the ranking of all alternatives at the same time therefore, it is easy to use by the DM.

It is assumed that the alternatives are evaluated by a weighted utility function that is defined as in (18).

$$v_j = \sum_{i=1}^{n} w_i x_{ij} \tag{18}$$

where x_{ij} denotes the evaluation of alternative j in criterion i and w_i denotes the weight of criterion i. Each criterion weight should be nonnegative, and all should sum up to 1.

At the first step of the proposed method, k random alternatives from the decision problem are presented to the DM and the DM is asked to rank them based on his/her preferences. Let the evaluation vector of alternative j on all criteria be represented by $\overline{x_j}$. If the resulting alternative ranking of DM is $\overline{x_k} \geq \overline{x_{k-1}} \geq \cdots \geq \overline{x_1}$, utility function values should also be ranked with the same order, i.e., $\overline{v_k} \geq \overline{v_{k-1}} \geq \cdots \geq \overline{v_1}$. However, there may be infinitely many criteria weights to satisfy these orders. To reach final answers, the authors proposed to find upper and lower bounds for each criterion weight. After determining the bounds, middlemost values are normalized to derive criteria weights. Upper bounds are found by solving the model given in (19).

$$\begin{gathered}
\max UB_i = w_i \\
\sum_{i=1}^{n} w_i x_{ij} = v_j\,, \quad j = 1, \ldots .. k \\
v_1 \leq v_2 \\
v_2 \leq v_3 \\
. \\
. \\
. \\
v_{k-1} \leq v_k \\
\sum_{i=1}^{n} w_i = 1 \\
w_i \geq 0, \quad i = 1, \ldots, n
\end{gathered} \tag{19}$$

By replacing the objective function with (20), lower bounds are calculated. Then, by taking the average of the lower and upper bounds for each criterion, average weights (AW_i) are obtained and normalized as in (21) to derive final criteria weights.

$$\min LB_i = w_i \tag{20}$$

$$w_i = \frac{AW_i}{\sum_{i=1}^{k} AW_i} \tag{21}$$

Alternatives are required to be ordered by the DM in clusters of k, but all rankings provided by DM are added to the same mathematical model to find final criteria weights.

3.2.2.3. SWING Method

SWING method proposed by Winterfeldt and Edwards (1986) is a another widely used weight elicitation approach. The method starts with generating *n* hypothetical alternatives for a problem with *n* criteria. Within these alternatives, each of them has the best value in one criterion and worst values in the other *n-1* criteria. The DM selects one of the *n* alternatives as the best one based on his/her preferences, then the criterion that has the best value in the selected alternative is set as the criterion with the highest preferability. In the next step, DM selects one of the remaining *n-1* alternatives, then the criterion with the best value on that alternative is determined as the second important criterion, and the procedure continues until the overall criteria ranking is derived.

After the criteria ranking is obtained, DM is asked to determine the value of each alternative relative to other alternatives to calculate weights. To do that, utility values of each alternative and utility of the nadir vector (vector with the worst values in all criteria) are compared. Afterwards, criteria weights are obtained by rating the differences between utility values of alternatives and the nadir vector.

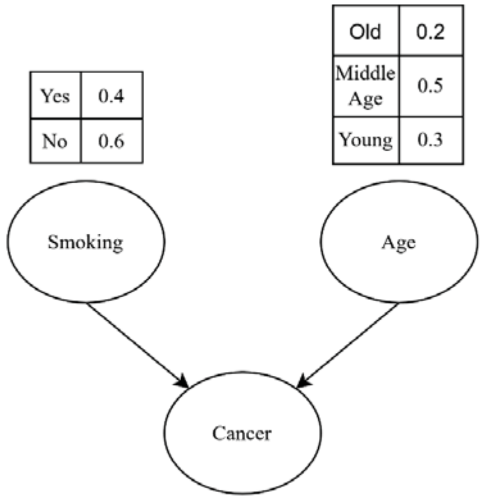
3.3. Bayesian Networks

BNs are probabilistic graphical models that represent joint probability distributions of variables. A BN model consists of a graphical structure and a probability table for each node in the graph. Its graphical structure is directed and acyclic, it includes nodes and arcs to represent variables and their relations, respectively. Since BNs show us which variables are independent, it eliminates the unnecessary conflicts from the model. BNs model the probability distributions, causal relationships and independencies graphically. Besides, they have powerful algorithms for probability calculations (Lauritzen and Spiegelhalter, 1988).

Using conditional probabilities, efficient algorithms of BNs can calculate posterior probabilities when an evidence is defined for some variables in the model. Also, it is possible to make probabilistic inferences with missing values. For more comprehensive information, Fenton and Neil (2012) may be reviewed.

To understand a BN model better, let us assume that smoking and age are two factors that affect the probability of having cancer. Smoking and age variables are called parent nodes while cancer variable is called child node because of its dependency on smoking and age. All relations between variables and their probability tables are shown in Figure 3.2. Probability tables on nodes are called node probability tables (NPT) and NPT of cancer variable is represented as conditional probability distributions of smoking and age variables.

BNs represent the joint probability distribution of its variables based on the conditional independencies encoded in the graphical structure and the conditional probability distributions encoded in NPTs. Propagation algorithms (Lauritzen and Spiegelhalter, 1988) can be used to compute the posterior distribution of its nodes when any subset of the nodes are instantiated. In the example case given in Figure 3.2, the probability of having cancer is 0.27 when none of the other nodes are instantiated. However, when an evidence is entered to the model, the posterior probabilities can be calculated. For example, if it is known that the patient is old then the model makes the probability of being cancer higher.



Age	Old	Old	Middle Age	Middle Age	Young	Young
Smoking	Yes	No	Yes	No	Yes	No
True	0.9	0.4	0.4	0.1	0.3	0.01
False	0.1	0.6	0.6	0.9	0.7	0.99

Figure 3.2. BN Model of the cancer example

Additionally, child nodes can be instantiated, and the probability distributions of parent nodes can be updated. For instance, if the patient has cancer, the probability of smoking increases from 0.4 to 0.7 or if the patient is young and has cancer, the probability of smoking is calculated as 0.95. BN models are effective tools to model joint probability distributions and make probabilistic inferences.

However, BNs are limited to calculation of posterior probabilities in decision support. Thus, to represent preferences of DMs in an MCDM problem, it needs another tool. In this thesis, we use BN models to gather probability distributions since we deal with uncertain alternative solutions on criteria.

4. PROPOSED METHODOLOGIES

Since preference functions, threshold values and criteria weights are DM specific, and there is uncertainty involved, it is not possible to assess the alternatives in a straightforward way. Thus, we need systematic and comprehensive approaches to compare alternatives and make decisions. In this thesis, we propose three different approaches to rank multiple criteria alternatives with uncertain evaluations. Our first and second approaches are the test-based and the score-based approach. The differences between them are the outranking information they use, the flexibility level they offer to the DM about the final evaluations of alternatives and their way of presenting results. The test-based approach uses PROMETHEE I rules to rank alternatives. It considers statistically significant positive and negative flow superiorities to perform a partial ranking. This approach enables the DM to define the error rate of their MCDM analysis by setting the significance level and power of the statistical test. On the other hand, the score-based approach uses PROMETHEE II to create a complete ranking of alternatives. It calculates the probabilities of alternatives for occupying different ranks. Based on these probabilities, this approach calculates an overall score to rank alternatives. It also enables the DM to assess the uncertainty in their ranking. Our third approach, a probabilistic version of PROMETHEE, works with joint probability distributions of alternative pairs. While the test-based and score-based approaches use sampling to deal with uncertain criteria values, the third approach uses probabilities to rank alternatives. In this section, we give explanations for all approaches.

In all approaches, importance of criteria is represented through weights. Different weight elicitation techniques were reviewed in Section 3.2. We mainly use AHP and ROC weighting to derive criteria weights in our approaches. In addition, we make additional computational studies for sensitivity analysis of weights. Although we propose the use of BNs as the source of uncertainty representation, our approaches can work with probability distributions generated by any means. In fact, since the BN of the shoulder case in the project does not have enough actual data to provide reliable distributions yet, we use expert knowledge to produce the probability distributions at this stage.

4.1. Test-Based Outranking Approach under Uncertainty

The test-based outranking approach uses statistical tests and confidence intervals to evaluate the differences between the means of both positive and negative flows of alternatives. For each pair of alternatives, firstly positive and negative means of alternatives and confidence intervals are calculated. To understand whether there is a significant difference between alternatives, statistical tests with significance level α are conducted.

Because we have uncertainty about criteria evaluations, we generate scenarios from probability distributions for criteria values. Let us call each of these scenarios a sample and let φ_{is}^{+} and φ_{is}^{-} be the positive and negative flow of solution a_i in sample s, respectively. Our test-based approach includes the steps given below:

1. For each sample s, apply PROMETHEE outranking steps 1-5 to calculate φ_{is}^{+} and φ_{is}^{-} values for all i
2. Check the mean of positive flows $\overline{\varphi_i^{+}}$ and negative flows $\overline{\varphi_i^{-}}$ for all i
   2.1. If the mean of positive flows $\overline{\varphi_i^{+}}$ and negative flows $\overline{\varphi_i^{-}}$ are the same for all i, conclude that there is indifference relationship between all solutions
   2.2. Else if there is any significant difference between $\overline{\varphi_i^{+}}$ or $\overline{\varphi_i^{-}}$
      2.2.1. For all solution pairs a_i and a_k, use φ_{is}^{+}, φ_{ks}^{+}, φ_{is}^{-} and φ_{ks}^{-} samples to test the following null-hypotheses H_0 with significance level α
      H_0^{Test1}: $\varphi_i^{+} = \varphi_k^{+}$
      H_0^{Test2}: $\varphi_i^{-} = \varphi_k^{-}$
      2.2.2. Conclude that
      There is an indifference relationship between a_i and a_k if:
         Fail to reject H_0^{Test1} and H_0^{Test2}
      Solution a_i is preferred to a_k if one of the following conditions hold:
         Reject H_0^{Test1} and H_0^{Test2}, $\overline{\varphi_i^{+}} > \overline{\varphi_k^{+}}$ and $\overline{\varphi_i^{-}} < \overline{\varphi_k^{-}}$
         Reject H_0^{Test1} and fail to reject H_0^{Test2}, $\overline{\varphi_i^{+}} > \overline{\varphi_k^{+}}$
         Fail to reject H_0^{Test1} and reject H_0^{Test2}, $\overline{\varphi_i^{-}} < \overline{\varphi_k^{-}}$
      Solutions a_i and a_k are incomparable if one of the following conditions hold:

Reject H_0 $^{\text{Test1}}$ and H_0 $^{\text{Test2}}$, $\overline{\varphi_i^+} > \overline{\varphi_k^+}$ and $\overline{\varphi_i^-} > \overline{\varphi_k^-}$

Reject H_0 $^{\text{Test1}}$ and H_0 $^{\text{Test2}}$, $\overline{\varphi_i^+} < \overline{\varphi_k^+}$ and $\overline{\varphi_i^-} < \overline{\varphi_k^-}$

3. Combine all relationships in Step 2 to construct the outranking diagram of all solutions

In Step 1, traditional PROMETHEE formulas (1) – (3) are performed for each sample and φ_{is}^+ and φ_{is}^- values for all i are calculated. After that, to reach an overall ranking of alternatives we apply statistical tests. In Step 2, mean of positive flows $\overline{\varphi_i^+}$and negative flows $\overline{\varphi_i^-}$ of all i are checked to understand whether the alternatives are significantly different or not. To do that, we use one-way Analysis of Variance (ANOVA) test. ANOVA test is applied for both positive and negative flow values separately. If ANOVA results say that mean of positive flows $\overline{\varphi_i^+}$ and negative flows $\overline{\varphi_i^-}$ are the same for all i, in Step 2.1 we conclude that alternatives are not significantly different and they all have indifference relationship. However, if ANOVA results show that positive $\overline{\varphi_i^+}$ or negative $\overline{\varphi_i^-}$ flow means have significant difference for at least one alternative, then we continue with Step 2.2 and make tests to identify alternative pairs which have significant difference. Therefore, in Step 2.2.1, we conduct two-sided Tukey's test for both positive and negative flows for all solution pairs a_i and a_k using φ_{is}^+, φ_{ks}^+, φ_{is}^- and φ_{ks}^-. Conducting multiple pairwise tests, such as paired t test, between the positive and negative flows of solutions is prone to false discoveries due to family-wise error rate. Tukey's test corrects for these errors when making pairwise comparison of the means. The tests are performed based on the following null hypotheses H_0 $^{\text{Test1}}$ and H_0 $^{\text{Test2}}$ with significance level α. Test 1 includes positive flows of alternatives in pairs while Test 2 includes negative flows.

H_0 $^{\text{Test1}}$: $\varphi_i^+ = \varphi_k^+$

H_0^{Test2}: $\varphi_i^- = \varphi_k^-$

After the Tukey's test is completed for both H_0 $^{\text{Test1}}$ and H_0 $^{\text{Test2}}$, conclusions are made based on the rules in Step 2.2.2. These rules are derived from the basic PROMETHEE rules, we updated them for the results of the Tukey's test. Finally, in Step 3 all relationships are aggregated, and ranking of alternatives are constructed.

As in traditional PROMETHEE, the test-based outranking approach generally gives partial ranking of alternatives since some alternatives have indifference relationship or they may be incomparable. This outcome may be preferable in some decision-making problems where DMs prefer a degree of flexibility in their final evaluation. For example, DMs may need a general assessment of alternative solutions, but they may not desire to force strict superiorities between the solutions. This property comes from the PROMETHEE I that considers positive and negative flows simultaneously. The classical PROMETHEE I, however, does not have a mechanism to define significant difference values between positive or negative flows of pairs of solutions when there are multiple evaluations with uncertainty. Our approach provides a guideline to make conclusions from data by taking the uncertainty and error rate into account by using statistical tests in PROMETHEE I.

The sample size may not be large enough to detect small differences between positive and negative flows in desired α levels in some cases. Conducting an automated analysis based on p-values may misguide the DM in assessing whether the difference was not present, or the sample size was too small to detect it. Therefore, a fully automated analysis is not recommended, and the DM should evaluate the confidence intervals of flow differences to assess whether large differences are included in the interval and their width.

4.2. Score-Based Approach under Uncertainty

Our second approach ranks alternatives based on the information about the probability of alternatives occupying each rank. Using these probabilities, an overall score is calculated in order to rank all alternatives. Additionally, this approach presents a graphical tool to evaluate performances of different solutions. The score-based approach uses PROMETHEE II rules to provide more exact results with complete ranking of alternatives. Thus, it works with net flows of alternatives. As in the first approach, to cope with the uncertainty in criteria values of alternatives, we utilize sampling approach for the probability distribution of criteria values. We define φ_{is} as the net flow of solution a_i in sample s. Based on φ_{is} values for all i in a given s, let r_{is} be the rank of solution a_i in sample s.

Steps of the proposed score-based approach follow:

1. For each sample s, apply PROMETHEE outranking steps 1-6 to calculate φ_{is} for all i.
2. For all s, list solutions in decreasing order of φ_{is} and determine r_{is} for all i accordingly.
3. Choose a suitable method for defining rank weights w_t.
4. Apply formulas (22) – (25) to compute $F_{r_i}(x)$ and θ_i for all i.
5. Draw a graphical summary of the ranking based on $F_{r_i}(x)$, and calculate the overall ranking of solutions by listing all solutions in decreasing order of θ_i.

For each alternative i in each sample s, classical PROMETHEE formulas (1)-(4) are calculated to find φ_{is} values in Step 1. Next, in Step 2, all i are listed in decreasing order of φ_{is} values for each s. Thus, r_{is} for all i are determined. In Step 3, we need to define weights to express the importance of ranks for the DM. The overall ranking scores are calculated with these rank weights to consider the risk behavior and preferences of the DM while providing the ranking of alternatives. If DM is unable or unwilling to determine weights directly, then one of the ranking-based weight elicitation methods discussed in Section 3.2.1 can be used. In our second approach, we prefer to use RS, RR or ROC weights using (5), (6) and (7). If certain ranks are considered to be not important, the DM may not consider them when defining rank weights and assign them 0 values. This should be done in accordance with the rule that weights should be non-increasing and sum to 1. With Step 4, we compute all calculations to find overall ranking scores of alternatives. Now, let S be the set of all samples, and S_{it} be the subset of S where a_i occupies the t^{th} rank as defined in (22).

$$S_{it} = \{s \in S | r_{is} = t\} \tag{22}$$

To find the probability of a_i having the t^{th} rank, p_{it}, we divide the cardinality of S_{it} by the cardinality of S as in (23).

$$p_{it} = \frac{|S_{it}|}{|S|} \tag{23}$$

p_{it} values only provide probabilities of all ranks for all alternatives, but we need an assessment tool to provide a complete ranking. For this purpose, we present a graphical tool and a score to aid the DM in ranking solutions. Our graphical tool includes cumulative distribution functions (CDFs) of all solutions for all possible ranks. Let r_i be the rank of solution a_i considering all $s \in S$; let the CDF of r_i, $F_{r_i}(x)$ to be defined as in (24).

$$F_{r_i}(x) = P(r_i \leq x) = \sum_{t=1}^{x} p_{it} \tag{24}$$

Creating a graph of these CDFs of all solutions gives us a summary of ranks of different solutions and the associated uncertainty. From the graph we can interpret some results like which solutions are better overall or which ones have higher probabilities to be in the top ranks.

In addition, we provide a final measure to rank solutions, ranking score θ_i which is calculated as in (25) and used to provide a complete ranking of n alternatives. w_t represents the weight of the t^{th} rank. This weighted score is flexible to decide based on the risk behavior and rank preferences of the DM.

$$\theta_i = \sum_{t=1}^{n} w_t p_{it} \tag{25}$$

To explain our approach in detail, let us construct an example with 4 solutions (a_1, a_2, a_3, a_4) and assume that we have 100 samples for criteria values of solutions. For this example, the number of criteria is not important since Table 4.1 includes the frequency of the rank occurrences and the related probabilities that are obtained with applying PROMETHEE II steps on each sample. For instance, a_1 has the first rank in 20 samples, the second rank in 10 samples, the third rank in 30 samples and the fourth rank in 40 samples. The columns on the right show the probabilities of occurrence of these ranks. For example, the probability of a_1 having the first rank is 0.2, the second rank is 0.1, the third rank is 0.3 and fourth rank is 0.4.

Table 4.1. Sample results of the toy example

Solution	$\lvert S_{i1}\rvert$	$\lvert S_{i2}\rvert$	$\lvert S_{i3}\rvert$	$\lvert S_{i4}\rvert$	p_{i1}	p_{i2}	p_{i3}	p_{i4}
a_1	20	10	30	40	0.2	0.1	0.3	0.4
a_2	10	20	40	30	0.1	0.2	0.4	0.3
a_3	40	20	20	20	0.4	0.2	0.2	0.2
a_4	30	50	10	10	0.3	0.5	0.1	0.1

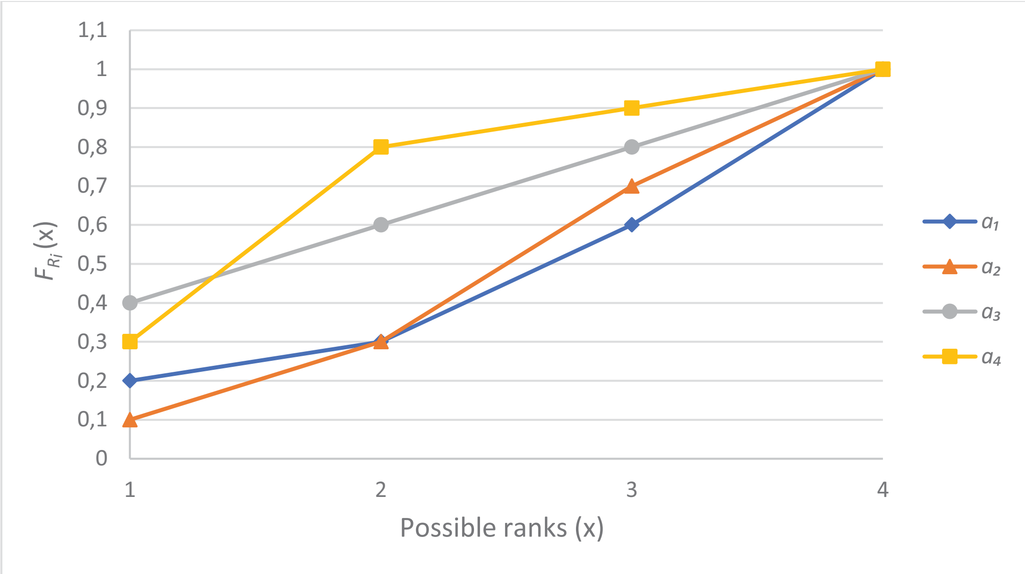


Figure 4.1. Cumulative probability plots of the example problem

Using (24), CDFs for the ranks of each solution are calculated and Figure 4.1 is constructed to show the graphs of those CDFs. In Table 4.1, we see that a_3 has the highest probability and a_2 has the lowest probability of having the first rank. Also, we can say that although a_3 has higher probability than a_4 for the first rank, a_4 may be considered as the better alternative for some DMs since it has higher probability for being the best or second-best solution. Thus, considering all ranks can be appropriate for the DMs who wants to evaluate all possible ranks. On the other hand, a_3 can be a better alternative than a_4 for the risk-seeking DMs since they can assign a relatively higher weight to the best rank.

Table 4.2 shows the weights (w_t) of each rank t calculated by different ranking approaches and weighted scores (θ_i) of alternative solutions using those weights. RS approach assumes equal distance between weights of different ranks; however, RR and ROC approaches may be more suitable weighting approaches for DMs who want to give considerably more importance to higher ranks. In this example, they nearly assign two times more weight to the first rank than the second one. In Table 4.2, the numbers in parentheses show the ranks of solutions. We also include a case where the DM uses RR weights for just the top two ranks. Using RS, a_4 has the first rank, a_3 has the second rank and a_1 and a_2 have the lowest ranks since they have equal weighted scores. With scores calculated by ROC, and RR for just the first 2 ranks, a_4 has the first rank, a_3 has the second rank, a_1 has the third rank and a_2 has the fourth. In the ranking by RR, however, the best alternative is a_3 which is followed by a_4 since RR puts more emphasis on the probability for the first rank relative to the weight of the second rank.

Table 4.2. Solution scores and rank weights of toy example

		Rank Weight Approach			
		RS	**ROC**	**RR**	**RR** *First 2 Ranks*
$\boldsymbol{w_t}$	w_1	0.40	0.52	0.48	0.66
	w_2	0.30	0.27	0.24	0.33
	w_3	0.20	0.15	0.16	0
	w_4	0.10	0.06	0.12	0
$\boldsymbol{\theta_i}$	θ_1	0.21 (3)	0.20 (3)	0.22 (3)	0.17 (3)
	θ_2	0.21 (3)	0.18 (4)	0.20 (4)	0.13 (4)
	θ_3	0.28 (2)	0.30 (2)	0.30 (1)	0.33 (2)
	θ_4	0.30 (1)	0.31 (1)	0.29 (2)	0.37 (1)

If the DM is risk-seeking and gives importance to just the top ranks, then all weighting methods considered can be used for only the top ranks and 0 weights can be assigned to other ranks as shown in the example. In general, if the DM is only interested in n ranks that corresponds to taking a more risk-seeking attitude than the case of considering all ranks, this would require changing N with n in (5) – (7), and assigning 0 weights to w_{n+1}, …, w_N.

The graphical tool given in Figure 4.1 is useful to observe differences between the performance of solutions. It shows that, for example, even though a_3 has a higher probability of having the first rank, a_4 has better overall probabilities when lower ranks are also considered. While the weighted score summarizes this in a single measure, the graph enables the DM to assess the differences in specific ranks graphically.

4.2.1. Analyzing the confidence of the score-based approach

Using the test-based approach proposed in Section 4.1, the confidence of the score-based approach can be assessed. Therefore, we can provide decision support with a confidence level to the DM. Recall that φ_{is} is the net flow of solution a_i in sample s. We can make tests to see whether the net flows of solutions are significantly different based on the ranking produced by the score-based approach. Let r_i^{score} be the rank of solution a_i provided by the score-based approach. Steps of the confidence assessment are as follows.

1. Test the presence of difference between the means of net flows $\overline{\varphi_i}$ of all i with ANOVA
2. If there is any significant difference
   2.1. For every solution pair a_i and a_k where the $r_i^{score} \leq r_k^{score}$
      2.1.1. Use φ_{is} and φ_{ks} samples to test the null-hypothesis H_0: $\varphi_i > \varphi_k$ with the significance level α using Tukey's Test.
3. Prepare a summary table with the p-values of the pairwise comparisons according to the ranking provided by the score-based approach.

This confidence assessment gives us a summary about the significant differences of solutions in the order of the ranking provided by the score-based approach. This assessment is advantageous since we cannot measure the insignificant difference between solutions when we work with net flow values of a sample in classical PROMETHEE II. Due to criteria weights, preference functions and threshold values, net flow values of a sample are very case specific. Thus, it is not straightforward to determine a cut-off value in net flows to determine classes of solutions that are significantly different from each other. However, by conducting statistical tests that we mentioned, we can test for significant differences between solutions. Therefore, the DM can be certain about the accuracy of the ranking with significance level α. We illustrate this analysis in computational experiments.

4.3. Probabilistic PROMETHEE

We also provide a third method of alternative assessment in addition to the previous test-based and score-based approaches. This third method uses joint probabilities instead of sampling approaches to provide rankings based on PROMETHEE methodology. We refer to this approach as the probabilistic PROMETHEE, and it works with joint probability distributions of alternatives in each criterion. It proceeds by taking the sum of probabilities where alternative i has better criterion values than alternative k for each alternative pair (*i,k*) and each criterion. Then, it calculates a weighted score for each alternative pair by using the importance weights of criteria to aggregate all scores. Finally, using the mechanisms of PROMETHEE, it provides negative and positive flows to perform both partial and complete alternative ranking as in the classical PROMETHEE approaches. Instead of the preference functions of PROMETHEE, this approach uses the probabilities of alternatives being better than the others.

Additionally, two different versions of this basic approach are proposed considering that the DM can be interested in giving different preferences to different magnitudes of difference in criteria values. The basic version of this approach gives equal importance to all differences in criterion values. However, the second version defines a threshold value to consider differences of alternatives as significant in accordance with the DM's preferences. For example, for an ordinal valued criterion, if the threshold value is 2 and alternative *i* is better than alternative *k* by 1 level in that criterion with a given probability, then the method will ignore that probability. Finally, the third version specifically gives different importance to each level of difference in criteria values. For instance, alternative *i* can be better than alternative *k* by 2 levels and alternative *l* can be better than alternative *k* by 1 level. In such a situation, the DM may prefer to treat alternatives *i* and *l* differently. Therefore, the third version of the probabilistic PROMETHEE approach defines a weighted probabilistic measure to rank alternatives.

Steps of the first version that we develop for ordinal-valued criteria are defined as follows:

1. Obtain joint probability J_{ik}^{j} distributions of all alternative pairs (i,k) in each criterion j
2. In each criterion j, use joint probability distributions to calculate dominance score P_{ik}^{j} of a_i on a_k as follows:
   2.1. For each criterion j where higher levels are preferable
   Sum the joint probabilities where a_i has higher levels than a_k
   2.2. For each criterion j where lower levels are preferable:
   Sum the joint probabilities where a_i has lower levels than a_k

3. Calculate weighted dominance scores M_{ik} of alternative pairs (i,k) regarding all criteria using (26) and create an outranking matrix with all dominance scores

$$M_{ik} = \sum_{j=1}^{n} \mathrm{w}_j P_{ik}^j \tag{26}$$

4. Find negative, positive and net outranking values of each alternative i using (27), (28) and (29)

$$\varphi_i^+ = \sum_k M_{ik} \tag{27}$$

$$\varphi_i^- = \sum_k M_{ki} \tag{28}$$

$$\varphi_i = \varphi_i^+ - \varphi_i^- \tag{29}$$

5. Obtain a partial or complete alternative ranking
   5.1. To obtain a partial ranking, compare negative φ_i^- and positive φ_i^+ outranking values and conclude that
   Alternatives a_i and a_k are indifferent if:
   $\varphi_i^+ = \varphi_k^+$ and $\varphi_i^- = \varphi_k^-$
   Alternative a_i is preferred to a_k if one of the following conditions hold:
   $\varphi_i^+ > \varphi_k^+$ and $\varphi_i^- < \varphi_k^-$
   $\varphi_i^+ > \varphi_k^+$ and $\varphi_i^- = \varphi_k^-$
   $\varphi_i^+ = \varphi_k^+$ and $\varphi_i^- < \varphi_k^-$
   Alternatives a_i and a_k are incomparable if one of the following conditions hold:
   $\varphi_i^+ > \varphi_k^+$ and $\varphi_i^- > \varphi_k^-$
   $\varphi_i^+ < \varphi_k^+$ and $\varphi_i^- < \varphi_k^-$
   5.2. To obtain a complete ranking, list alternatives in decreasing order of φ_i values

Let us use an example to explain the probabilistic PROMETHEE approach fully. Let a_1, a_2 and a_3 be three solution alternatives evaluated with two ordinal criteria and assume that the first criterion has three levels and the second one has four levels. Also, assume that higher levels of both criteria are preferable for the DM. For this example, weights of criterion 1 and 2 are taken as 0.4 and 0.6, respectively. Table 4.3, Table 4.4 and Table 4.5 report some of the joint probability distributions of alternative pairs.

Table 4.3. Joint probability distribution of alternatives 1 and 2 in the first criterion

		a_2		
		Low	**Medium**	**High**
a_1	**Low**	0.07	0.04	0.09
	Medium	0.13	0.08	0.18
	High	0.14	0.08	0.20

Table 4.4. Joint probability distribution of alternatives 1 and 2 in the second criterion

		a_2			
		Very Low	**Low**	**Medium**	**High**
a_1	**Very Low**	0.04	0.03	0.08	0.07
	Low	0.05	0.03	0.09	0.08
	Medium	0.02	0.01	0.04	0.04
	High	0.08	0.05	0.14	0.14

Table 4.5. Joint probability distribution of alternatives 2 and 3 in the second criterion

		a_3			
		Very Low	**Low**	**Medium**	**High**
a_2	**Very Low**	0.02	0.06	0.03	0.09
	Low	0.01	0.03	0.02	0.05
	Medium	0.04	0.10	0.05	0.16
	High	0.04	0.10	0.05	0.15

In Step 2, using joint probability distributions, dominance scores (P_{ik}^{j}) are calculated. In our example, some P_{ik}^{j} scores using Table 4.3, Table 4.4 and Table 4.5 are calculated below as examples.

$$P_{12}^{1} = 0.13 + 0.14 + 0.08 = 0.35$$

$$P_{21}^{1} = 0.04 + 0.09 + 0.18 = 0.31$$

$$P_{21}^{2} = 0.03 + 0.08 + 0.07 + 0.09 + 0.08 + 0.04 = 0.39$$

$$P_{23}^{2} = 0.01 + 0.04 + 0.10 + 0.04 + 0.10 + 0.05 = 0.34$$

In Step 3, the weighted dominance scores are found by weighing the P_{ik}^{j} scores by criteria weights. For instance, the weighted dominance score M_{21} of a_2 over a_1 is found as below and the dominance matrix regarding weighted dominance scores of each alternative pair is provided in Table 4.6.

$$M_{21} = (P_{21}^1 * w_1) + (P_{21}^2 * w_2) = 0.31 * 0.4 + 0.39 * 0.6 = 0.36$$

Table 4.6. Dominance matrix

	a_1	a_2	a_3
a_1	0.00	0.35	0.34
a_2	0.36	0.00	0.35
a_3	0.34	0.37	0.00

Positive, negative and net outranking values of each alternative are calculated in Step 4 and reported in Table 4.7. According to these results, in both partial and complete rankings a_3 outranks a_1 and a_2. In partial ranking, a_1 and a_2 are found as incomparable, and in the complete ranking they are both placed on the second rank since their net outranking values are equal.

Table 4.7. Probabilistic positive, negative and net outranking values of alternatives

	φ^+	φ^-	φ
a_1	0.69	0.70	-0.01
a_2	0.71	0.72	-0.01
a_3	0.71	0.68	0.02

4.3.1. The Second Version of Probabilistic PROMETHEE: Defining Threshold Values

In this version of the probabilistic PROMETHEE, a threshold value t^j is defined for each desired criterion j according to the preferences of the DM. While dominance scores P_{ik}^j are calculated, probabilities where alternatives have a difference of at least t^j levels are included. This second version is similar to the basic version except for Step 2.

Let us explain this version using the same example. Assume that the threshold value is 2 for both criteria. Thus, the joint probabilities where alternatives have at least 2 level differences on both criteria are considered. From Table 4.3, dominance score of a_1 over a_2 on the first criterion, P_{12}^1 ,is found as 0.14 since just the probability where a_1 is on the high and a_2 is on the low level of the first criterion is considered. Similarly, other dominance scores are updated below.

$$P_{21}^{1} = 0.09$$

$$P_{21}^{2} = 0.08 + 0.07 + 0.08 = 0.23$$

$$P_{23}^{2} = 0.04 + 0.04 + 0.10 = 0.18$$

Again, the probabilities where a_2 has medium or high values when a_3 has very low values and the probabilities where a_2 has high values when a_3 has low values on criterion 2 are considered. Domination scores of each alternative pair are calculated similarly. All other steps are same with the first version. Table 4.8 shows the dominance scores of each alternative and Table 4.9 shows the outranking values of alternatives. In both partial and complete ranking, a_2 is the best alternative and followed by a_3 and a_1, respectively.

Table 4.8. Dominance matrix calculated by the second version of probabilistic PROMETHEE

	a_1	a_2	a_3
a_1	0.00	0.15	0.17
a_2	0.17	0.00	0.17
a_3	0.18	0.15	0.00

Table 4.9. Positive, negative and net outranking values of alternatives of the second version of probabilistic PROMETHEE

	φ^+	φ^-	φ
a_1	0.31	0.35	-0.04
a_2	0.35	0.30	0.05
a_3	0.33	0.34	-0.01

4.3.2. The Third Version of Probabilistic PROMETHEE: Defining Weighted Probabilistic Scores

In this version, the aim is to differentiate between each level of difference in criteria. The motivation is that, it may not be fair to evaluate alternative *i,* which is 3 levels better than alternative *j,* in the same way as alternative *l,* which is better than alternative *j* by 2 levels. Thus, we propose a weighted probabilistic score WP_{ik}^{j} instead of dominance score P_{ik}^{j} using (30). Let s_j be the number of levels of criterion j and J_{ik}^{jd} is the sum of joint probabilities of alternatives i and k in criterion j where the criterion level difference between i and k is d. After WP_{ik}^{j} scores are calculated, weighted dominance scores M_{ik} are obtained using (31). The remaining steps of the process are similar to the previous versions.

$$WP_{ik}^{j} = \frac{\sum_{d=1}^{s_j-1} J_{ik}^{jd} d}{\sum_{d=1}^{s_j-1} d} \tag{30}$$

$$M_{ik} = \sum_{j=1}^{n} WP_{ik}^{j} \mathrm{w}_j \tag{31}$$

In the same example, based on the probabilities in Table 4.3, weighted probabilistic score WP_{12}^{1} of a_1 over a_2 in criterion 1 is calculated as below. Since criterion 1 has 3 levels ($s_1=3$), criterion level differences (d) can be 1 or 2.

$$WP_{12}^{1} = \frac{(0.13 + 0.08) * 1 + 0.14 * 2}{1 + 2} = \frac{0.49}{1 + 2} = 0.16$$

Similarly, since criterion 2 has 4 levels ($s_2=4$), the following formula gives the weighted probabilistic score WP_{23}^{2} of a_2 over a_3 in criterion 2.

$$WP_{23}^{2} = \frac{(0.01 + 0.10 + 0.05) * 1 + (0.04 + 0.10) * 2 + 0.04 * 3}{1 + 2 + 3} = \frac{0.56}{1 + 2 + 3} = 0.09$$

Then, the same steps with the previous versions are applied with those new weighted scores in order to construct the dominance matrix and hence, the final ranking. Table 4.10 and Table 4.11 show the results. Both partial and complete rankings give the same ranking: a_1 is the best alternative and a_3 is the worst alternative.

Table 4.10. Dominance matrix calculated by the third version of probabilistic PROMETHEE

	a_1	a_2	a_3
a_1	0.00	0.12	0.13
a_2	0.12	0.00	0.13
a_3	0.12	0.12	0.00

Table 4.11. Positive, negative and net outranking values of alternatives of the third version of probabilistic PROMETHEE

	φ^+	φ^-	φ
a_1	0.25	0.24	0.01
a_2	0.25	0.25	0.00
a_3	0.24	0.25	-0.01

In the basic version, a_3 dominates the other two alternatives since all differences in criteria values have equal importance. However, in the second version of the approach, when at least 2 level difference is considered for criteria levels of alternatives, a_2 achieves the first rank, thus we conclude that a_2 has higher probabilities than a_3 with at least 2 level difference. This version can be applied where the DM wants to ignore insignificant differences completely. Lastly, in the third version, a_1 outranks both a_2 and a_3 since all differences are weighted in accordance with the magnitude of the differences. When we ignore the differences, which are less than 2, a_2 outranks a_1 but when we include lower differences with lower weights as in the third version, a_1 outranks a_2. Thus, we can conclude that, if a DM who gives importance to any level of difference between alternatives wants to highlight the alternatives which have higher probabilities for higher differences, then the third version will be useful for decision support.

5. COMPUTATIONAL STUDIES

To illustrate our test-based, score-based and probabilistic PROMETHEE approaches, we use two different cases. One of them is a treatment selection case for a patient who is suffering from shoulder pain. This case is used to assess the success of our approaches in a medical problem. Since we have conducted this thesis as a part of a TUBITAK project, we utilize the expertise of physiotherapists in the project to create the shoulder pain case. The proposed approaches are agnostic to how probability distributions of criteria are generated, hence they can work with probabilities generated by domain experts or quantitative modelling approaches such as BNs. The first case study uses probability distributions elicited from expert physiotherapists. The second case study is conducted to test our approaches in an area apart from healthcare. It uses a BN model to compute the posterior probability of criteria. We use the problem proposed by Kaya and Yet (2019) to present another approach to a supplier selection problem. For this case, probability distributions are obtained from a BN model generated by Kaya and Yet (2019).

For both treatment selection for shoulder pain and supplier selection cases, test-based, score-based and probabilistic PROMETHEE approaches are tested.

5.1. Treatment Selection Case for Shoulder Pain

To test our approaches on a medical decision-making problem, we use a shoulder pain case created by physiotherapists. A 32-year-old male is suffering from a recurrent shoulder problem. He is a painter and decorator, but he is struggling to continue his job, so he is worried about having a rest. According to his condition, eight treatment alternatives are considered, these are NSAIDs (non-steroidal anti-inflammatory drugs), using web-based advice sheets, physio rehab, injection, injection with physio rehab, waiting and seeing, surgery, and physio rehab with surgery. These treatments are evaluated with eight criteria: discomfort caused by treatment, improvement in functional ability, improvement in pain, psychosocial improvement, recovery time, side effects, time spent for treatment and waiting time to receive treatment. Also, there is another criterion that can be considered in a similar problem called quality adjusted life year (QALY). It is included in medical problems recently, but we cannot use that criteria in our case since it is not possible to generate probability distributions for QALY with experts. Since the treatment results cannot be known before the treatments are applied, we work with probabilistic values on criteria. In the case of using a BN model for this case, the only requirement to obtain probability distributions would be to enter patient-based data such as age, gender, test and evaluation results and medical history to the BN model. Then, the BN model would calculate the probability distributions of alternatives in each criterion. In our application, we derive probability distributions of alternatives in each criterion based on the expert information given by physiotherapists. Table 5.1 shows probability distributions of treatments in each criterion. Levels of waiting time, time spent and recovery time criteria are derived from the time intervals provided by the experts. Levels of the remaining criteria are supplied by the experts according to their usual clinical practice. We represent levels of all criteria with linguistic terms for simplicity. All criteria have different number of levels that can come from N–None, VL–Very Low, L–Low, L-M – Low-Medium, M–Medium, H–High, VH–Very High. The levels with asterisk symbols represent the best level of each criterion. Preference functions and corresponding threshold values are reported in Table 5.3. The thresholds represent the required level of difference for strict preference in the criteria with Type III preference function. For instance, since the threshold value of side effects criterion is two, at least a two level difference is required for strict preference for this criterion such as N vs. M, L vs. H, or N vs. H. Any level of difference on criteria with Type I function is accepted to cause strict preference. We determined preference functions and threshold values with the experts considering sensible levels of differences in criteria that would induce partial and strict preference. These functions can be easily updated for different clinicians and patient types. To follow treatment names easily, abbreviations in Table 5.2 are used in the rest of the thesis.

Table 5.1. Probability distributions of each treatment on each criterion

	Improvement in Functional Ability			Improvement in Pain			Discomfort			Psychosocial Improvement		
Treatment Alternatives	H*	M	L	H*	M	L	L*	M	H	H*	M	L
NSAIDs	0.399	0.577	0.024	0.226	0.601	0.173	0.630	0.370	0.000	0.000	0.370	0.630
Web Advice Sheet	0.028	0.598	0.374	0.352	0.567	0.081	1.000	0.000	0.000	0.005	0.315	0.680
Physio Rehab	0.630	0.370	0.000	0.450	0.532	0.018	0.326	0.636	0.038	0.622	0.358	0.021
Injection	0.000	0.447	0.553	0.794	0.205	0.001	0.226	0.601	0.173	0.149	0.594	0.256
Injection and Physio Rehab	0.696	0.301	0.003	0.826	0.173	0.001	0.326	0.636	0.038	0.757	0.241	0.002
Wait and See	0.092	0.553	0.355	0.004	0.744	0.252	0.748	0.252	0.000	0.001	0.193	0.806
Surgery and Physio Rehab	0.702	0.298	0.000	0.001	0.205	0.794	0.000	0.204	0.796	0.334	0.618	0.048
Surgery	0.000	0.447	0.553	0.002	0.280	0.718	0.002	0.260	0.738	0.091	0.647	0.262

	Recovery Time						Waiting Time					
Treatment Alternatives	VL*	L	L-M	M	H	VH	VL*	L	L-M	M	H	VH
NSAIDs	0.289	0.698	0.013	0.000	0.000	0.000	0.041	0.432	0.472	0.054	0.001	0.000
Web Advice Sheet	0.187	0.784	0.029	0.000	0.000	0.000	0.907	0.093	0.000	0.000	0.000	0.000
Physio Rehab	0.668	0.331	0.001	0.000	0.000	0.000	0.000	0.003	0.224	0.682	0.091	0.000
Injection	0.001	0.328	0.662	0.010	0.000	0.000	0.000	0.011	0.391	0.565	0.033	0.000
Injection and Physio Rehab	0.607	0.392	0.001	0.000	0.000	0.000	0.000	0.000	0.004	0.290	0.644	0.062
Wait and See	0.189	0.647	0.161	0.002	0.000	0.000	0.334	0.618	0.048	0.000	0.000	0.000
Surgery and Physio Rehab	0.000	0.000	0.000	0.289	0.698	0.013	0.000	0.000	0.000	0.000	0.140	0.860
Surgery	0.000	0.000	0.002	0.629	0.369	0.000	0.000	0.000	0.000	0.000	0.223	0.777

	Side Effects				Time Spent					
Treatment Alternatives	N*	L	M	H	VL*	L	L-M	M	H	VH
NSAIDs	0.019	0.172	0.453	0.356	0.001	0.389	0.603	0.006	0.000	0.000
Web Advice Sheet	0.830	0.170	0.000	0.000	0.607	0.392	0.001	0.000	0.000	0.000
Physio Rehab	0.380	0.606	0.014	0.000	0.000	0.001	0.369	0.623	0.007	0.000
Injection	0.123	0.655	0.218	0.004	0.289	0.698	0.013	0.000	0.000	0.000
Injection and Physio Rehab	0.308	0.670	0.022	0.000	0.000	0.000	0.067	0.653	0.276	0.004
Wait and See	0.840	0.158	0.001	0.000	0.028	0.539	0.422	0.012	0.000	0.000
Surgery and Physio Rehab	0.000	0.003	0.285	0.712	0.000	0.000	0.000	0.048	0.618	0.334
Surgery	0.000	0.001	0.170	0.830	0.000	0.000	0.000	0.001	0.392	0.607

Table 5.2. Treatment representations

Treatment Alternatives	Representations
NSAIDs	T1
Web Advice Sheet	T2
Physio Rehab	T3
Injection	T4
Injection with Physio Rehab	T5
Wait and See	T6
Surgery with Physio Rehab	T7
Surgery	T8

Table 5.3. Preference functions and thresholds of criteria for treatment selection

Criteria	Discomfort	Func. Ability	Imp. in Pain	Psychosocial Imp.	Recovery Time	Side Effects	Time Spent	Waiting Time
Preference Function	I	I	I	III	III	III	III	I
Threshold Value	-	-	-	*p=2*	*p=3*	*p=2*	*p=3*	-

Since this treatment selection problem has patient-based solutions, criteria have different weights for different patients. Also, it is repeated several times in a day for different patients, so an easy weight elicitation method is more appropriate than a method that requires several pairwise comparisons like AHP. Therefore, we used ROC approach to derive weights in this case and they are confirmed by the domain experts. Criteria weights are reported in Table 5.4. 100 samples are generated from probabilities given in Table 5.1. The following sections report the results of the test-based, score-based and probabilistic PROMETHEE approaches for the shoulder pain case.

Table 5.4. Criteria weights derived by ROC for treatment selection

Criteria	Discomfort	Functional Ability	Imp. in Pain	Psychosoc Imp.	Recovery Time	Side Effects	Time Spent	Waiting Time
Weights	0.111	0.340	0.152	0.079	0.033	0.215	0.016	0.054

Results of our approaches can be used by both clinicians and patients. If the criteria ranking is determined by the collaboration of the clinician and the patient, approaches are expected to provide more satisfactory results due to the advantages of shared decision making.

5.1.1. Results of the Test-based Outranking Approach for Shoulder Pain

Following the steps given in Section 4.1, φ_{is}^{+} and φ_{is}^{-} values are calculated and ANOVA test is conducted to find out if means of positive flows $\overline{\varphi_i^{+}}$ and negative flows $\overline{\varphi_i^{-}}$ are significantly different or not for at least one treatment. Table 5.5 reports the results of ANOVA tests for both positive and negative flows of treatments. Since the *p*-values are less than the significance level $\alpha=0.05$, we reject both null hypotheses below.

H_0^+: $\overline{\varphi_1^+} = \overline{\varphi_2^+} = \overline{\varphi_3^+} = \overline{\varphi_4^+} = \overline{\varphi_5^+} = \overline{\varphi_6^+} = \overline{\varphi_7^+} = \overline{\varphi_8^+}$
H_0^-: $\overline{\varphi_1^-} = \overline{\varphi_2^-} = \overline{\varphi_3^-} = \overline{\varphi_4^-} = \overline{\varphi_5^-} = \overline{\varphi_6^-} = \overline{\varphi_7^-} = \overline{\varphi_8^-}$

Thus, we can say that at least one treatment exists whose mean positive flow and mean negative flow are significantly different from those of other treatments.

Table 5.5. ANOVA results for means of positive and negative flows for treatment selection

	Source	Df	Sum Sq.	Mean Sq.	F Value	P Value
Positive flows	Treatments	7	663.6	94.8	232.5	$<2e^{-16}$
	Residuals	792	323.0	0.4		
Negative flows	Treatments	7	954.1	136.3	305.7	$<2e^{-16}$
	Residuals	792	353.2	0.5		

In order to identify which treatment pairs are significantly different, we continue with post-hoc analysis. Tukey's Test with %95 confidence level is conducted for the following null hypotheses for all treatment pairs.

H_0^{Test1}: $\varphi_i^+ = \varphi_k^+$
H_0^{Test2}: $\varphi_i^- = \varphi_k^-$

The results for positive and negative flows are reported in Table 5.6 and Table 5.7, respectively. Tables show the mean difference between treatment pairs and corresponding confidence intervals. The differences are calculated with the treatment in the row minus the treatment in the column and the asterisk symbol represents that the corresponding p-value is less than 0.05. Analyzing confidence intervals, treatment pairs which are not significantly different can be observed, for example T2-T1 cell in Table 5.6 shows that although mean positive flow of Treatment 2 is higher than that of Treatment 1 by 0.14, it is not a significant difference since the confidence interval includes 0 and thus, p-value is greater than 0.05. So, we fail to reject the null hypothesis that H_0^{Test1}: $\varphi_1^+ = \varphi_2^+$. Also, we can see that mean positive flow of Treatment 3 is significantly higher than that of Treatment 1 in T3-T1 cell in Table 5.6. The same cell in Table 5.7 shows that the mean negative flow of Treatment 3 is significantly lower than that of Treatment 1 this time, thus we conclude that Treatment 3 outranks Treatment 1.

Table 5.6. Tukey's Test results for positive flows of treatments

	T1	T2	T3	T4	T5	T6	T7	T8
T1	-							
T2	(-0.14;0.41) 0.14	-						
T3	(0.66;1.21) 0.94*	(0.53;1.08) 0.80*	-					
T4	(-0.66;-0.11) -0.39*	(-0.80;-0.25) -0.52*	(-1.60;-1.05) -1.32*	-				
T5	(0.83;1.38) 1.10*	(0.69;1.24) 0.97*	(-0.11;0.44) 0.17	(1.22;1.77) 1.49*	-			
T6	(-0.67;-0.12) -0.39*	(-0.80;-0.25) -0.53*	(-1.61;-1.06) -1.33*	(-0.28;0.27) -0.01	(-1.77;-1.22) -1.50*	-		
T7	(-1.03;-0.48) -0.76*	(-1.17;-0.62) -0.90*	(-1.97;-1.42) -1.70*	(-0.65;-0.10) -0.37*	(-2.14;-1.59) -1.86*	(-0.64;-0.09) -0.37*	-	
T8	(-2.24;-1.69) -1.96*	(-2.37;-1.83) -2.10*	(-3.18;-2.63) -2.90*	(-1.85;-1.30) -1.58*	(-3.34;-2.79) -3.07*	(-1.85;-1.30) -1.57*	(-1.48;-0.93) -1.20*	-

Table 5.7. Tukey's Test results for negative flows of treatments

	T1	T2	T3	T4	T5	T6	T7	T8
T1	-							
T2	(-0.25;0.33) 0.04	-						
T3	(-1.07;-0.49) -0.78*	(-1.11;-0.53) -0.82*	-					
T4	(0.30;0.87) 0.59*	(0.26;0.83) 0.55*	(1.08;1.65) 1.37*	-				
T5	(-1.12;-0.55) -0.84*	(-1.16;-0.59) -0.87*	(-0.34;0.23) -0.05	(-1.71;-1.13) -1.42*	-			
T6	(0.19;0.76) 0.47*	(0.15;0.72) 0.44*	(0.97;1.54) 1.26*	(-0.40;0.18) -0.11	(1.02;1.60) 1.31*	-		
T7	(1.15;1.73) 1.44*	(1.11;1.69) 1.40*	(1.93;2.51) 2.22*	(0.57;1.14) 0.85*	(1.99;2.56) 2.27*	(0.68;1.25) 0.97*	-	
T8	(2.40;2.97) 2.69*	(2.36;2.94) 2.65*	(3.18;3.76) 3.47*	(1.81;2.39) 2.10*	(3.23;3.81) 3.52*	(1.93;2.50) 2.21*	(0.96;1.53) 1.25*	-

According to the test-based approach, the outranking relations of treatments are shown in Figure 5.1. According to the results, the best treatment alternatives for this patient are Treatment 3 and 5 (Physio Rehab and Injection with Physio Rehab) and the worst one is

Treatment 8 (Surgery). There is indifference between some treatment pairs: Treatment 3 and 5, Treatment 1 and 2, and Treatment 4 and 6. Physiotherapists may select Treatment 5 or 3 as the best treatment based on the preference of the patient. Treatment 1 or 2 are placed on the second-best rank. They are followed by Treatment 4 and 6.

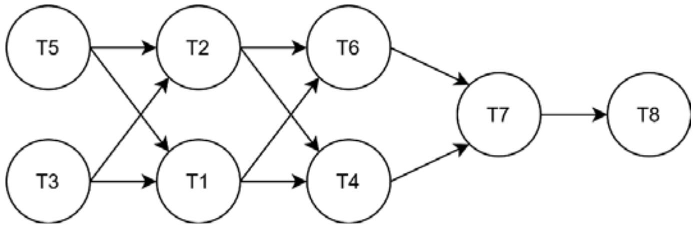


Figure 5.1. Outranking relation of treatments with the test-based approach

5.1.2. Results of the Score-based Outranking Approach for Shoulder Pain

To apply the score-based approach to the shoulder pain case, the steps mentioned in Section 4.2 are followed. After φ_{is} values are calculated for each treatment and sample *s,* treatment rankings are generated in the descending order of φ_{is} values for each *s.* To prioritize ranks, three weighting methods are applied: RS, ROC and RR. The detailed explanations of these methods are discussed in Section 3.2.1.2. Table 5.8 shows the rank weights required to calculate weighted scores θ_i of treatments. The last three columns include the weights that are derived considering only the first three ranks.

Table 5.8. Rank weights for RS, ROC and RR approaches for treatments

Rank (t)	**W_t(RS)**	**W_t(ROC)**	**W_t(RR)**	**W_t(RS)** *first 3 ranks*	**W_t(ROC)** *first 3 ranks*	**W_t(RR)** *first 3 ranks*
1	0.222	0.340	0.368	0.500	0.611	0.545
2	0.194	0.215	0.184	0.333	0.278	0.273
3	0.167	0.152	0.123	0.167	0.111	0.182
4	0.139	0.111	0.092	0.000	0.000	0.000
5	0.111	0.079	0.074	0.000	0.000	0.000
6	0.083	0.054	0.061	0.000	0.000	0.000
7	0.056	0.033	0.053	0.000	0.000	0.000
8	0.028	0.016	0.046	0.000	0.000	0.000

Then, probabilities of each treatment occupying each rank are calculated and reported in Table 5.9. Also, the graphical summary of treatment rankings with cumulative distribution $F_{r_i}(x)$ is shown in Figure 5.2.

Table 5.9. Probabilities of treatments for occupying each rank

Rank	T1	T2	T3	T4	T5	T6	T7	T8
1	0.110	0.040	0.380	0.020	0.440	0.010	0.000	0.000
2	0.100	0.140	0.320	0.040	0.360	0.040	0.000	0.000
3	0.140	0.330	0.140	0.160	0.110	0.120	0.000	0.000
4	0.270	0.210	0.100	0.150	0.060	0.190	0.020	0.000
5	0.190	0.180	0.040	0.170	0.020	0.270	0.110	0.020
6	0.110	0.030	0.020	0.300	0.010	0.200	0.320	0.010
7	0.080	0.070	0.000	0.140	0.000	0.160	0.490	0.060
8	0.000	0.000	0.000	0.020	0.000	0.010	0.060	0.910

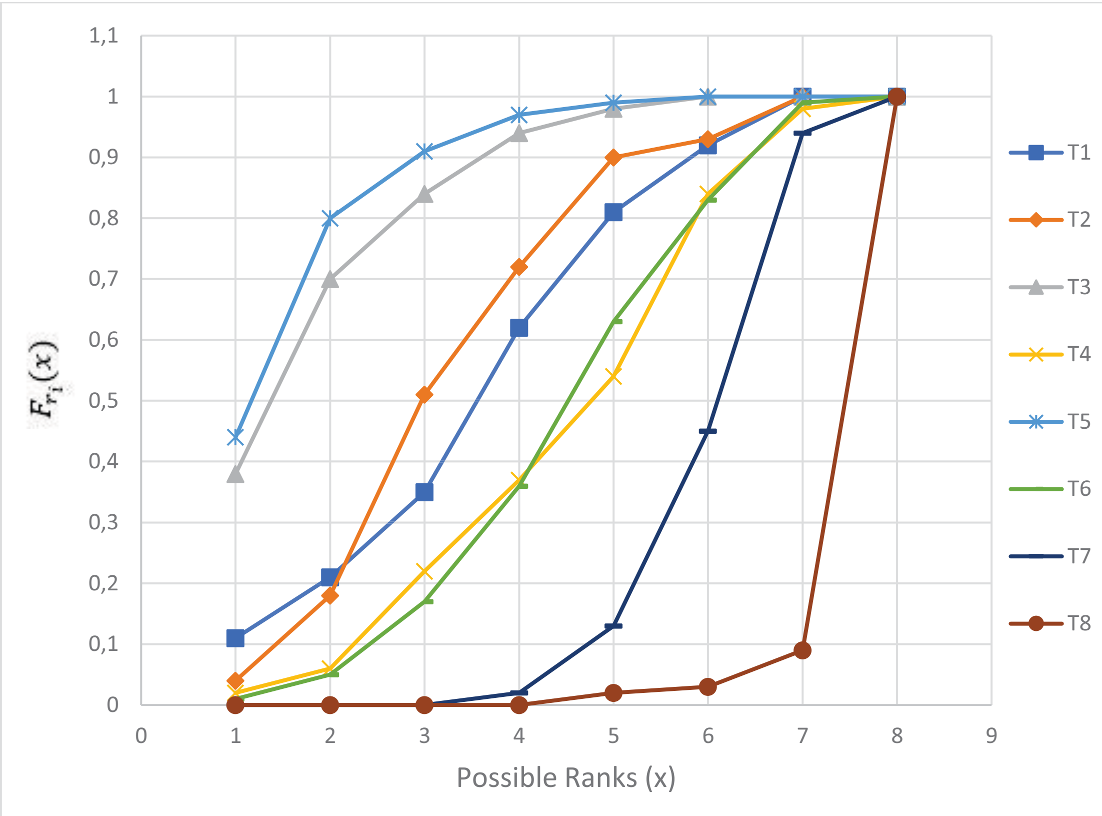


Figure 5.2. Cumulative probability plots of treatments

The complete treatment ranking obtained using ROC weights for ranks is illustrated in Figure 5.3. Treatment 5 (Injection with Physio Rehab) is the best alternative, followed by Treatment 3 (Physio Rehab). In Table 5.9, it is seen that these two treatments have the highest probabilities to occupy the first rank. However, although Treatment 1 has higher probability than Treatment 2 for the first rank, Treatment 2 outranks Treatment 1 in overall since Treatment 2 achieves higher probabilities for second and third ranks. Analyzing Figure 5.2, it is seen that Treatment 4 and 6 are close to each other. Until the fourth rank, Treatment 4 has higher probabilities, then Treatment 6 closes this difference and performs better in lower ranks. However, since the first ranks have higher weights, Treatment 4 outranks Treatment 6. The best and worst alternatives are the same with the result of the test-based approach. The treatments with indifference relationship in the test-based approach such as Treatments 5 and 3, Treatments 2 and 1, and Treatments 4 and 6 are now fully ranked in the score-based approach.

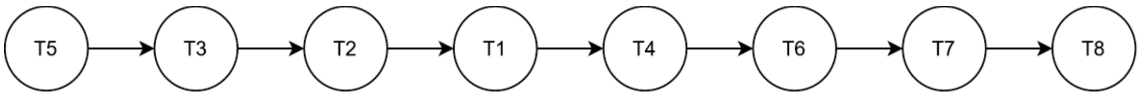


Figure 5.3. Outranking relations of treatments in the score-based approach using ROC weights

Weighted scores and rankings of treatments according to the three rank weighting approaches are reported in Table 5.10 and Table 5.11. In Table 5.10, it is assumed that the DM considers all possible ranks, but in Table 5.11, it is assumed that the DM considers only the first three ranks.

Table 5.10. θ_i scores and final rankings of treatments when all ranks are considered

Rank	Treatment	θ_i (RS)	Treatment	θ_i (ROC)	Treatment	θ_i (RR)
1	T5	0.1975	T5	0.2523	T5	0.2492
2	T3	0.1900	T3	0.2344	T3	0.2292
3	T2	0.1467	T2	0.1354	T1	0.1258
4	T1	0.1394	T1	0.1337	T2	0.1190
5	T6	0.1122	T4	0.0911	T4	0.0873
6	T4	0.1119	T6	0.0891	T6	0.0842
7	T7	0.0706	T7	0.0457	T7	0.0581
8	T8	0.0317	T8	0.0184	T8	0.0471

If the DM is risk-averse and wants to consider the performances of treatments in all ranks, then results in Table 5.10 can be used. For all weighting approaches, Treatments 5 (Injection with Physio Rehab) and 3 (Physio Rehab) are the two best treatments and they are followed by Treatments 2 and 1. In the rankings of both RS and ROC approaches, Treatment 2 outranks Treatment 1. However, in the ranking of RR, Treatment 1 outranks Treatment 2 since the weight of the first rank is much higher than the weight of second rank in RR and Treatment 1 has higher probability for the first rank. On the other hand, in both ROC and RR approaches this time, Treatment 4 outranks Treatment 6, but in RS approach, Treatment 6 outranks Treatment 4. The reason is that RS assumes equal distances between all consecutive weights while ROC and RR approaches assume larger differences between first ranks. Treatment 4 has higher probabilities than Treatment 6 for the first ranks so, Treatment 4 outranks Treatment 6 using RR and ROC approaches but cannot outrank using RS. Additionally, the worst alternatives, Treatments 7 and 8 are the same for all weighting approaches.

Table 5.11. θ_i scores and final rankings of treatments when the first three ranks are considered

Rank	Treatment	θ_i (RS)	Treatment	θ_i (ROC)	Treatment	θ_i (RR)
1	T5	0.3583	T5	0.3811	T5	0.3582
2	T3	0.3200	T3	0.3367	T3	0.3200
3	T2	0.1217	T1	0.1106	T2	0.1200
4	T1	0.1117	T2	0.1000	T1	0.1127
5	T4	0.0500	T4	0.0411	T4	0.0509
6	T6	0.0383	T6	0.0306	T6	0.0382
7	T7	0.0000	T7	0.0000	T7	0.0000
8	T8	0.0000	T8	0.0000	T8	0.0000

On the other hand, for a risk-seeking DM, Table 5.11 can be used since it calculates the weighted scores of treatments considering just the first three ranks. Again, for all weighting approaches, Treatment 5 and 3 are the best options. They are followed by Treatments 2 and 1 in RS and RR approaches while Treatment 1 outranks Treatment 2 in ROC approach. When we analyze Figure 5.2, we see that Treatment 1 has higher probability for the first rank but Treatment 2 closes the difference and outranks Treatment 1 when they come to the third rank since Treatment 2 has higher probabilities for the second and the third ranks. In ROC approach, Treatment 1 outranks Treatment 2 with a small difference since weight of the first rank is much higher than the second and the third ranks. Meanwhile, in RS and RR approaches, the reverse is obtained since the differences between first three ranks are much lower than ROC. Finally, the last four treatments are the same for all weighting approaches.

As discussed in Section 4.2.1, to test the confidence of the treatment ranking provided by the score-based approach in Figure 5.3, ANOVA test and Tukey's Test are conducted. ANOVA results in Table 5.12 shows that at least one treatment in the ranking is significantly different from other treatments. To test the significance of difference between consecutive treatments in the ranking list, results in Table 5.13 are analyzed.

Table 5.12. ANOVA results for means of net flows for treatment selection

Source	Df	Sum Sq.	Mean Sq.	F Value	P Value
Treatments	7	3185	455	331.2	<2e-16
Residuals	792	1088	1.4		

Table 5.13. Tukey's Test results for net flows of treatment pairs

	T1	T2	T3	T4	T5	T6	T7	T8
T1	-							
T2	(-0.41;0.60) 0.10	-						
T3	(1.22;2.22) 1.72*	(1.12;2.13) 1.62*	-					
T4	(-1.48;-0.47) -0.97*	(-1.57;-0.57) -1.07*	(-3.20;-2.19) -2.69*	-				
T5	(1.44;2.44) 1.94*	(1.34;2.35) 1.84*	(-0.28;0.72) 0.22	(2.41;3.42) 2.91*	-			
T6	(-1.37;-0.36) -0.87*	(-1.47;-0.46) -0.97*	(-3.09;-2.08) -2.59*	(-0.40;0.61) 0.10	(-3.31;-2.30) -2.81*	-		
T7	(-2.70;-1.70) -2.20*	(-2.80;-1.79) -2.30*	(-4.42;-3.41) -3.92*	(-1.73;-0.72) -1.23*	(-4.64;-3.63) -4.14*	(-1.83;-0.83) -1.33*	-	
T8	(-5.15;-4.15) -4.65*	(-5.25;-4.24) -4.75*	(-6.87;-5.87) -6.37*	(-4.18;-3.17) -3.68*	(-7.09;-6.09) -6.59*	(-4.29;-3.28) -3.78*	(-2.96;-1.95) -2.45*	-

Tukey's Test results in Table 5.13 show that some consecutive treatment alternatives in the ranking provided in Figure 5.3 are not significantly different such as Treatments 1 and 2, Treatments 3 and 5, and Treatments 4 and 6. With the score-based approach, a complete ranking is provided for the DM. But if the DM does not necessarily require a complete ranking, some insignificant outranking relations may be eliminated using the confidence test.

5.1.2.1. Sensitivity Analysis on Criteria Weights for Shoulder Pain

Because the output of our approaches, as in many MCDM methods, depends on the weights of criteria, we also make sensitivity analysis to see how our results are affected with changes in weights and whether they are robust to small changes. For the score-based approach, we conduct sensitivity analysis to determine allowable ranges for criteria weights for the results to remain stable. We use the weight stability intervals procedure by Mareschal (1988) for this task by modifying it to be applicable with uncertain criteria evaluations. The weight stability intervals procedure is developed to find intervals for criteria weights so that the given ranks of solutions according to additive utility functions do not change, and it can be applied with PROMETHEE II scores. It uses the differences in criteria values between successive solutions in the rank list to find these intervals. The intervals represent the changes that can be made to a single criterion while the others are kept constant (but normalized to ensure the summation of weights is still one). Interested readers can consult Mareschal (1988) for the details of the procedure. We can only make use of it for the score-based approach since our other approaches do not use scores compatible with this analysis. In the next case study of supplier selection, we make further sensitivity analysis for all of our approaches by changing the AHP weights with ROC weights and comparing the results.

The weight stability intervals procedure can only work with certain criteria evaluations in a single sample, so we need to enhance it to work under uncertainty in the score-based approach. Using the procedure in Mareschal (1988), we construct the interval of each criterion weight in each sample. Each sample produces its separate ranking of solutions, so we arrive at 100 intervals for each criterion weight. Now these intervals need to be aggregated into an overall interval for each criterion, but this is not straightforward. Taking the tightest interval among all samples is not suitable as some intervals can be very narrow, and some can even consist of only the original weight. As a result, we propose to form the overall intervals of the weights with the values that appear in at least a given percentage of all samples. Since all the intervals in the samples are formed around the original weights, these original weights appear in 100% of the intervals. As we move away from the original weights, the percentage of samples that contain the value in consideration gets smaller. In line with the logic of confidence intervals, we use 95% as the cut-off value. Taking ROC weights given in Table 5.4 as the original weights, Table 5.14 reports the resulting aggregated intervals we obtain. We argue that, as long as the weights are changed within these intervals, we can be 95% confident that the ranking list of PROMETHEE II will not change. This in turn will ensure that the scores of the score-based approach will not change. The results in Table 5.14 suggest the weights of functional ability and improvement in pain should be set carefully since they have relatively narrow ranges. On the other hand, the ranking list is not so sensitive to changes in the weights of other criteria, so uncertainties in those areas can be tolerated better.

Table 5.14. Weights stability intervals for ROC weights

Criteria	ROC Weight	Weight Stability Interval
Discomfort	0.111	[0.096 – 0.236]
Functional Ability	0.340	[0.301 – 0.375]
Improvement in Pain	0.152	[0.129 – 0.272]
Psychosocial Improvement	0.079	[0.000 – 0.240]
Recovery Time	0.033	[0.000 – 0.394]
Side Effects	0.215	[0.000 – 0.228]
Time Spent	0.016	[0.000 – 0.292]
Waiting Time	0.054	[0.000 – 0.188]

5.1.3. Results of Probabilistic PROMETHEE Approach for Shoulder Pain

Using the rules explained in Section 4.3, we perform three experiments regarding three versions of the approach. The first version of the approach provides the outranking values in Table 5.15. These relations result in the same ranking of treatments for both PROMETHEE I and II rules; there are no indifference relations between treatment pairs when we consider positive and negative flows simultaneously. The resulting ranking is shown in Figure 5.4.

Table 5.15. Outranking values by the first version of probabilistic PROMETHEE

		φ^+	φ^-	φ
T1	NSAIDs	2.468	2.360	0.108
T2	Web Advice Sheet	2.964	1.904	1.060
T3	Physio Rehab	3.594	1.242	2.352
T4	Injection	2.204	2.824	-0.620
T5	Injection with Physio	3.816	1.164	2.652
T6	Wait and See	2.625	2.227	0.398
T7	Surgery with Physio	1.844	3.491	-1.647
T8	Surgery	0.517	4.820	-4.303

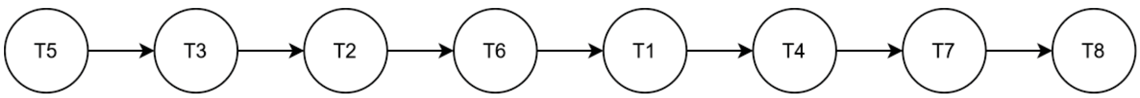


Figure 5.4. Ranking of treatments by the first version of probabilistic PROMETHEE

Treatment 5 is the best option for the patient, followed by Treatment 3 and 2, whereas the worst one is again Treatment 8.

In the second version of the problem, we determine the threshold value as 2 for side effects, recovery time and time spent for treatment. It means that in these criteria, at least 2 level difference between treatments is considered. Otherwise, the joint probability of these treatment pairs is assumed as 0. Table 5.16 lists the positive, negative and net flows of each treatment and Figure 5.5 illustrates the complete ranking of treatments which is again the same for PROMETHEE I and II rules. Treatments 5, 3 and 2 are the best three alternatives for the second version too, and the positions of the worst treatments, Treatments 4, 7 and 8, are similar to the first version as well. The only difference of the second version from the first version of the probabilistic PROMETHEE is the relation between Treatments 1 and 6. This time, Treatment 1 outranks Treatment 6. Therefore, we conclude that Treatment 6 is better than Treatment 1 with small differences in the first version of the approach, since when a threshold value of 2 is defined in the second version, Treatment 1 provides better results than Treatment 6.

Table 5.16. Outranking values by the second version of probabilistic PROMETHEE

		φ^+	φ^-	φ
T1	NSAIDs	2.220	1.933	0.287
T2	Web Advice Sheet	2.455	1.793	0.662
T3	Physio Rehab	3.199	0.923	2.276
T4	Injection	1.889	2.314	-0.425
T5	Injection with Physio	3.443	0.802	2.641
T6	Wait and See	2.118	2.101	0.017
T7	Surgery with Physio	1.760	3.165	-1.405
T8	Surgery	0.461	4.514	-4.053

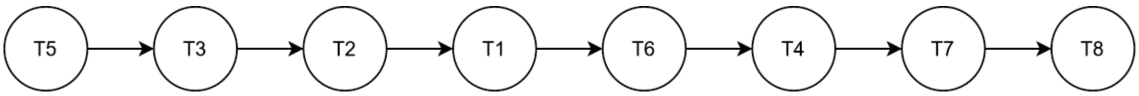


Figure 5.5. Ranking of treatments by the second version of probabilistic PROMETHEE

Table 5.17 provides outranking values of each treatment alternative performed by the third version of the probabilistic PROMETHEE, and these values are used to create partial and complete rankings as shown in Figure 5.6 and Figure 5.7, respectively.

Table 5.17. Outranking values by the third version of probabilistic PROMETHEE

		φ^{+}	φ^{-}	φ
T1	NSAIDs	0.862	0.778	0.084
T2	Web Advice Sheet	0.996	0.762	0.234
T3	Physio Rehab	1.344	0.324	1.020
T4	Injection	0.739	0.999	-0.260
T5	Injection with Physio	1.493	0.279	1.214
T6	Wait and See	0.847	0.876	-0.029
T7	Surgery with Physio	0.769	1.293	-0.524
T8	Surgery	0.163	1.903	-1.740

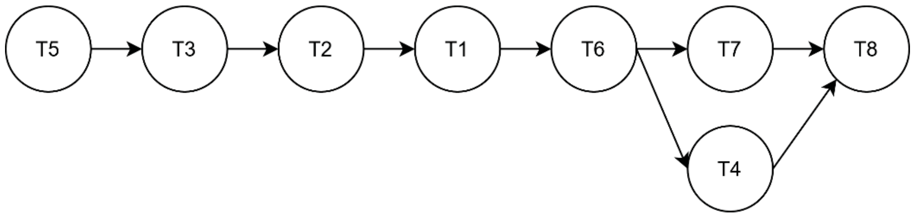


Figure 5.6. Partial ranking of treatments by the third version of probabilistic PROMETHEE

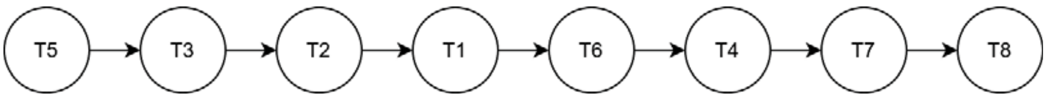


Figure 5.7. Complete ranking of treatments by the third version of probabilistic PROMETHEE

It is seen that the rankings in Figure 5.6 and Figure 5.7 do not conflict but the partial ranking shown in Figure 5.6 differs by the incomparability relation of Treatments 7 and 4. All other outranking relations are mutual for both partial and complete ranking. Treatments 5, 3, and 2 are still the best alternatives as in previous approaches. Treatment 8 is still the worst alternative. When a complete ranking is required by the DM, Figure 5.7 can be used. On the other hand, based on the preferences of the DM or the patient, Treatment 7 can be considered instead of Treatment 4 using Figure 5.6.

If we compare the third version of the approach with other versions, we can report that the best three treatments and the worst treatment are the same. Also, when higher weights are defined for higher level differences on criteria, complete ranking gives the same result with the second version of the approach. For instance, in the results of both versions, Treatment 5 performs better than Treatment 3 and Treatment 1 has better results than Treatment 6 with higher level differences. Thus, it can be deduced that outranking relations in the second version of the approach are already caused by high differences in criteria.

Our results are validated by the physiotherapists in the project. The resulting rankings are compatible with their experiences and expectations. They also expressed that our methods offer the advantages of summarizing the performance of alternatives in different criteria and considering preferences about different criteria explicitly.

5.2. Supplier Selection Case

Since supplier selection is a widely studied MCDM problem in the literature (see Chai, Liu and Ngai, 2013; Govindan et al., 2015; Zimmer, Fröhling and Schultmann, 2016 for reviews of decision analysis and MCDM models in supplier selection) we prefer to apply our approaches in this area too. Besides, this case has an available BN. The BN model of Kaya and Yet (2019) evaluates suppliers based on seven criteria: product quality, cost, delivery performance, quality system certifications, flexibility, cooperation, and reputation. Cost and quality system certifications criteria can be observed before working with suppliers; thus, they have deterministic values on these criteria. On the other hand, the other five criteria cannot be known with certainty beforehand, so suppliers have probabilistic values on these five criteria and the BN can estimate the related probabilities based on past data and evidence. All criteria have five ordinal states: VL-very low, L-low, M – medium, H – high and VH – very high. We use the BN model to generate data for ten different suppliers. Table 5.18 shows the probability distributions of each criterion for each supplier.

Table 5.18. Probability distributions of the suppliers in each criterion

	Product Quality					Cost					Delivery Performance					Quality System Certifications				
	VL	L	M	H	VH	VL	L	M	H	VH	VL	L	M	H	VH	VL	L	M	H	VH
S1	0.000	0.001	0.068	0.825	0.106	0.000	0.000	0.000	1.000	0.000	0.001	0.029	0.679	0.291	0.000	0.000	0.000	0.000	1.000	0.000
S2	0.000	0.001	0.189	0.682	0.128	0.000	0.000	1.000	0.000	0.000	0.020	0.164	0.419	0.322	0.075	0.000	0.000	0.000	0.000	1.000
S3	0.000	0.001	0.013	0.626	0.360	0.000	0.000	0.000	1.000	0.000	0.000	0.025	0.601	0.374	0.000	0.000	0.000	0.000	1.000	0.000
S4	0.000	0.001	0.280	0.706	0.013	0.000	0.000	0.000	1.000	0.000	0.000	0.007	0.132	0.487	0.374	0.000	0.000	1.000	0.000	0.000
S5	0.000	0.000	0.788	0.207	0.005	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.015	0.899	0.086	0.000	0.000	1.000	0.000	0.000
S6	0.001	0.003	0.336	0.610	0.050	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.003	0.673	0.324	0.000	1.000	0.000	0.000	0.000
S7	0.000	0.001	0.026	0.628	0.345	0.000	0.000	0.000	1.000	0.000	0.010	0.864	0.126	0.000	0.000	0.000	1.000	0.000	0.000	0.000
S8	0.000	0.028	0.972	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.005	0.822	0.173	0.000	0.000	0.000	0.000	1.000
S9	0.000	0.000	0.437	0.558	0.005	0.000	0.000	0.000	1.000	0.000	0.002	0.064	0.390	0.443	0.101	0.000	0.000	0.000	1.000	0.000
S10	0.000	0.000	0.999	0.001	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.004	0.126	0.530	0.340	0.000	0.000	0.000	1.000	0.000

	Flexibility					Cooperation					Reputation				
	VL	L	M	H	VH	VL	L	M	H	VH	VL	L	M	H	VH
S1	0.186	0.801	0.013	0.000	0.000	0.000	0.006	0.994	0.000	0.000	0.008	0.328	0.663	0.001	0.000
S2	0.027	0.127	0.295	0.342	0.209	0.000	0.000	0.108	0.820	0.072	0.000	0.000	0.997	0.003	0.000
S3	0.012	0.085	0.266	0.379	0.258	0.000	0.001	0.078	0.831	0.090	0.001	0.102	0.896	0.001	0.000
S4	0.001	0.096	0.655	0.248	0.000	0.000	0.001	0.002	0.323	0.674	0.000	0.000	0.999	0.001	0.000
S5	0.226	0.766	0.008	0.000	0.000	0.002	0.106	0.562	0.317	0.013	0.001	0.015	0.141	0.433	0.411
S6	0.000	0.000	0.002	0.475	0.523	0.001	0.014	0.844	0.141	0.000	0.011	0.368	0.621	0.000	0.000
S7	0.038	0.205	0.399	0.282	0.076	0.001	0.001	0.253	0.708	0.037	0.000	0.000	0.999	0.000	0.000
S8	0.000	0.076	0.622	0.302	0.000	0.000	0.000	0.034	0.519	0.447	0.002	0.154	0.843	0.001	0.000
S9	0.190	0.796	0.014	0.000	0.000	0.000	0.000	0.478	0.501	0.021	0.000	0.000	0.998	0.002	0.000
S10	0.176	0.809	0.015	0.000	0.000	0.000	0.000	0.477	0.503	0.020	0.010	0.370	0.619	0.001	0.000

Preference functions of criteria, threshold values and criteria weights are determined with the domain expert in supplier selection in Kaya and Yet (2019). Table 5.19 shows the functions and threshold values. Since the threshold value of flexibility criteria is two, at least a two level difference is required for strict preference for this criterion such as VL vs. M, L vs. H, M vs. VH, VL vs. H, L vs. VH, or VL vs. VH. For the quality system certificates criterion, only two or higher level differences are accepted for strict preference. For Type I preference functions, any level of difference is enough for strict preference between two solutions.

Table 5.19. Preference functions and thresholds of criteria

Criteria	Product Quality	Cost	Delivery Perf.	Quality Syst. Cert.	Flexibility	Cooperation	Reputation
Preference Function Type	I	I	III	II	III	III	I
Threshold Value	-	-	*p*=3	*q*=1	*p*=2	*p*=2	-

In this case, criteria weights are determined once, so we have the opportunity to use a more detailed approach for weight elicitation. Criteria weights are determined using AHP based on the comparisons of the domain expert. The resulting weights are given in Table 5.20. The consistency of comparisons is found as 0.051. In Section 5.2.4, we also check the results with ROC weights.

Table 5.20. Criteria weights by AHP

Criteria	Product Quality	Cost	Delivery Perf.	Quality Syst. Cert.	Flexibility	Cooperation	Reputation
Weights	0.356	0.231	0.230	0.075	0.032	0.042	0.034

100 random samples are generated from the BN model. The samples represent the data available for criteria evaluations of suppliers. The following sections present the results of the test-based, score-based and probabilistic PROMETHEE approaches.

5.2.1. Results of the Test-based Outranking Approach for Supplier Selection

Results in Table 5.21 show that there is at least one supplier whose positive and negative flows are both significantly different from other suppliers since the null hypotheses H_0: $\overline{\varphi_1^+}=$

$\overline{\varphi_2^+}= \overline{\varphi_3^+}= \overline{\varphi_4^+}= \overline{\varphi_5^+}=\overline{\varphi_6^+}= \overline{\varphi_7^+}= \overline{\varphi_8^+}= \overline{\varphi_9^+}= \overline{\varphi_{10}^+}$ and H_0: $\overline{\varphi_1^-}= \overline{\varphi_2^-}= \overline{\varphi_3^-}= \overline{\varphi_4^-}= \overline{\varphi_5^-}=\overline{\varphi_6^-}= \overline{\varphi_7^-}= \overline{\varphi_8^-}= \overline{\varphi_9^-}= \overline{\varphi_{10}^-}$ are rejected because *p*-values are less than the significance level α= 0.05.

Table 5.21. ANOVA results for the means of positive and negative flows of suppliers

	Source	Df	Sum Sq.	Mean Sq.	F Value	P Value
Positive flows	Suppliers	9	461.1	51.24	88.18	$<2e^{-16}$
	Residuals	990	575.3	0.58		
Negative flows	Suppliers	9	340.9	37.87	81.51	$<2e^{-16}$
	Residuals	990	460.0	0.46		

In the next step, to identify specifically which supplier pairs are significantly different, Tukey's test with %95 confidence level is conducted for the following null hypotheses for all supplier pairs (*i,k*).

H_0^{Test1}: $\varphi_i^+ = \varphi_k^+$
H_0^{Test2}: $\varphi_i^- = \varphi_k^-$

Table 5.22 and Table 5.23 show the results of the test for positive and negative flows of pairs, respectively. As in the shoulder pain case, the tables report the mean difference between pairs and the corresponding confidence intervals, also asterisk symbols represent that the *p*-value is less than 0.05. We can see null hypothesis of which supplier pair is rejected. For example, S2 – S1 cell of Table 5.22 reports that mean positive flow of S2 is significantly higher than the mean positive flow of S1. Also, the same cell in Table 5.23 shows that the mean negative flow of S2 is significantly lower than the mean negative flow of S1. Thus, we conclude that S2 outranks S1. Analyzing the results of all supplier pairs in Table 5.22 and Table 5.23 using the same approach discussed in the previous case study, a relation diagram is constructed as in Figure 5.8.

Table 5.22. Tukey's Test results for positive flows of supplier pairs

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
S1	-									
S2	(0.98;1.67) 1.33*	-								
S3	(0.31;1.00) 0.65*	(-1.01;-0.33) -0.67*	-							
S4	(-0.37;0.31) -0.03	(-1.70;-1.01) -1.35*	(-1.02;-0.34) -0.68*	-						
S5	(-0.30;0.39) 0.05	(-1.62;-0.94) -1.28*	(-0.95;-0.27) -0.61*	(-0.27;0.42) 0.07	-					
S6	(1.26;1.94) 1.60*	(-0.07;0.62) 0.28	(0.61;1.29) 0.95*	(1.29;1.97) 1.63*	(1.21;1.90) 1.56*	-				
S7	(-0.20;0.48) 0.14	(-1.53;-0.84) -1.19*	(-0.85;-0.17) -0.51*	(-0.17;0.51) 0.17	(-0.25;0.44) 0.09	(-1.80;-1.12) -1.46*	-			
S8	(0.98;1.66) 1.32*	(-0.35;0.33) -0.01	(0.32;1.01) 0.67*	(1.01;1.69) 1.35*	(0.93;1.61) 1.27*	(-0.62;0.06) -0.28	(0.84;1.52) 1.18*	-		
S9	(-0.52;0.17) -0.18	(-1.84;-1.16) -1.50*	(-1.17;-0.49) -0.83*	(-0.49;0.19) -0.15	(-0.56;0.12) -0.22	(-2.12;-1.44) -1.78*	(-0.66;0.03) -0.32	(-1.84;-1.15) -1.49*	-	
S10	(-0.69;-0.01) -0.35*	(-2.02;-1.33) -1.68*	(-1.35;-0.66) -1.00*	(-0.66;0.02) -0.32	(-0.74;-0.05) -0.40*	(-2.29;-1.61) -1.95*	(-0.83;-0.15) -0.49*	(-2.01;-1.33) -1.67*	(-0.52;0.17) -0.17	-

Table 5.23. Tukey's Test results for negative flows of supplier pairs

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
S1	-									
S2	(-1.16;-0.55) -0.85*	-								
S3	(-0.81;-0.20) -0.51*	(0.04;0.65) 0.35*	-							
S4	(-0.30;0.32) 0.01	(0.56;1.17) 0.86*	(0.21;0.82) 0.52*	-						
S5	(-0.12;0.49) 0.18	(0.73;1.34) 1.04*	(0.38;0.99) 0.69*	(-0.13;0.48) 0.17	-					
S6	(-1.04;-0.42) -0.73*	(-0.18;0.43) 0.12	(-0.53;0.08) -0.22	(-1.05;-0.43) -0.74*	(-1.22;-0.61) -0.91*	-				
S7	(0.69;1.30) 0.99*	(1.54;2.15) 1.85*	(1.19;1.81) 1.50*	(0.68;1.29) 0.98*	(0.51;1.12) 0.81*	(1.42;2.03) 1.72*	-			
S8	(-0.22;0.39) 0.09	(0.63;1.25) 0.94*	(0.29;0.90) 0.59*	(-0.23;0.38) 0.08	(-0.40;0.21) -0.10	(0.51;1.12) 0.82*	(-1.21;-0.60) -0.91*	-		
S9	(-0.18;0.43) 0.12	(0.67;1.28) 0.98*	(0.32;0.93) 0.63*	(-0.19;0.42) 0.11	(-0.37;0.24) -0.06	(0.55;1.16) 0.85*	(-1.18;-0.57) -0.87*	(-0.27;0.34) 0.04*	-	
S10	(0.62;1.23) 0.92*	(1.47;2.08) 1.78*	(1.13;1.74) 1.43*	(0.61;1.22) 0.91*	(0.44;1.05) 0.74*	(1.35;1.96) 1.65*	(-0.37;0.24) -0.07	(0.53;1.14) 0.84*	(0.50;1.11) 0.80*	-

According to the results in Figure 5.8, Supplier 2 and Supplier 6 are the best alternatives since they are not outranked by any supplier. They are followed by Supplier 3 and 8. Suppliers 1, 4, 5 and 9 are the next group of alternatives. They are followed by Supplier 7 and then the worst alternative, Supplier 10.

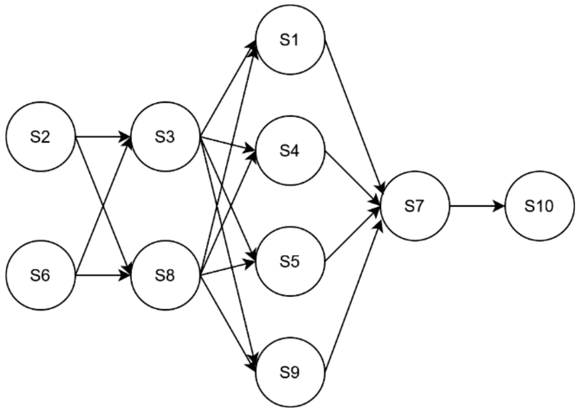


Figure 5.8. Outranking relation of suppliers for test-based outranking approach

5.2.2. Results of the Score-based Outranking Approach for Supplier Selection

After φ_{is} values are calculated for each alternative i and sample s, supplier rankings are generated in the descending order of φ_{is} values for each s. Table 5.24 reports the rank weights derived by RS, ROC and RR approaches. The last three columns show the weights when just first three ranks are considered by the DM.

Table 5.24. Rank weights derived by RS, ROC and RR for supplier selection

Rank (t)	**W_t(RS)**	**W_t(ROC)**	**W_t(RR)**	**W_t(RS)** *first 3 ranks*	**W_t(ROC)** *first 3 ranks*	**W_t(RR)** *first 3 ranks*
1	0.182	0.293	0.341	0.500	0.611	0.545
2	0.164	0.193	0.171	0.333	0.278	0.273
3	0.145	0.143	0.114	0.167	0.111	0.182
4	0.127	0.110	0.085	0.000	0.000	0.000
5	0.109	0.085	0.068	0.000	0.000	0.000
6	0.091	0.065	0.057	0.000	0.000	0.000
7	0.073	0.048	0.049	0.000	0.000	0.000
8	0.055	0.034	0.043	0.000	0.000	0.000
9	0.036	0.021	0.038	0.000	0.000	0.000
10	0.018	0.010	0.034	0.000	0.000	0.000

In the next step, $F_{r_i}(x)$ and θ_i for each supplier i are calculated. Table 5.25 reports the rank probabilities of suppliers and Figure 5.9 shows a graphical summary of supplier rankings based on cumulative distributions $F_{r_i}(x)$.

Table 5.25. Probabilities of the suppliers occupying each rank

Rank	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
1	0.000	0.400	0.060	0.020	0.030	0.480	0.000	0.010	0.000	0.000
2	0.050	0.230	0.230	0.060	0.060	0.200	0.000	0.150	0.020	0.000
3	0.070	0.080	0.170	0.090	0.110	0.030	0.040	0.320	0.090	0.000
4	0.090	0.020	0.120	0.120	0.040	0.040	0.050	0.350	0.170	0.000
5	0.090	0.050	0.170	0.190	0.010	0.070	0.150	0.080	0.180	0.010
6	0.300	0.080	0.120	0.110	0.050	0.040	0.070	0.080	0.110	0.040
7	0.170	0.040	0.070	0.080	0.250	0.030	0.130	0.000	0.080	0.150
8	0.130	0.050	0.040	0.070	0.250	0.090	0.120	0.010	0.060	0.180
9	0.060	0.030	0.010	0.060	0.130	0.020	0.260	0.000	0.070	0.360
10	0.040	0.020	0.010	0.200	0.070	0.000	0.180	0.000	0.220	0.260

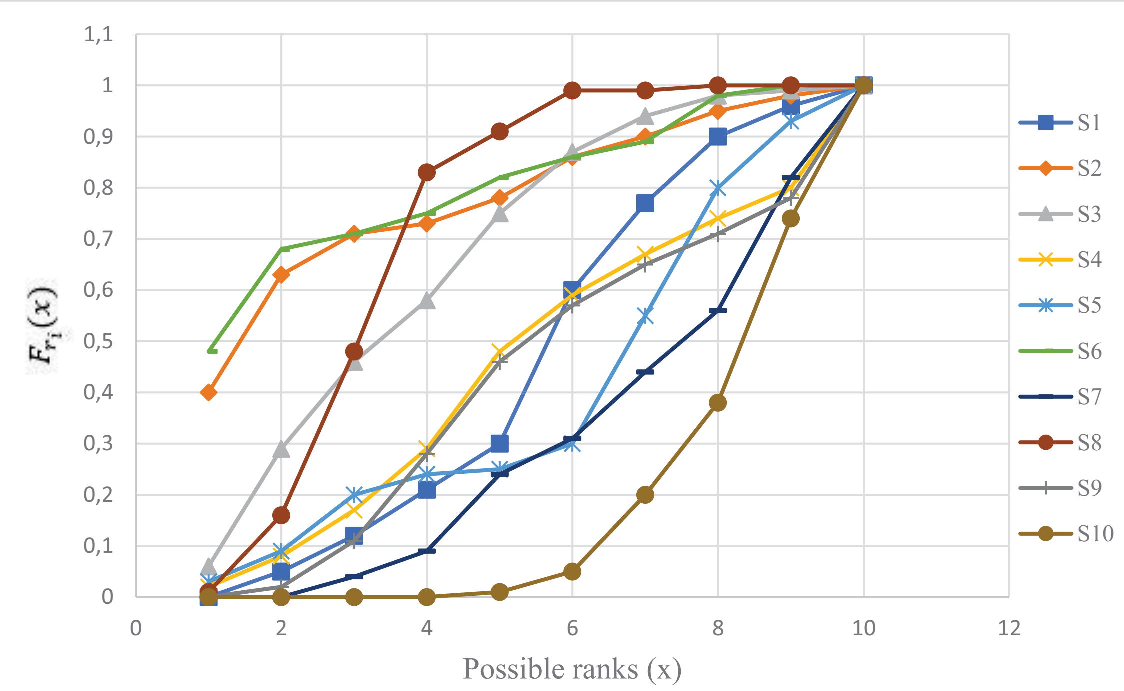


Figure 5.9. Cumulative probability plots of suppliers

Figure 5.10 shows the outranking relation of suppliers based on weighted scores θ_i when rank weights are derived by ROC approach. It is seen that Supplier 6 is the best alternative, followed by Supplier 2, and they are the suppliers which have the highest probabilities for the first rank. However, although Supplier 8 has the sixth highest probability for the first rank, it achieves the third rank in overall since its probabilities on the third and fourth ranks are much higher than the others. Also, when Suppliers 4 and 5 are compared, we can observe that Supplier 5 has higher probabilities in higher ranks, but Supplier 4 has better probabilities in the remaining ranks. Thus, in overall Supplier 4 outranks Supplier 5. In Figure 5.9, it is seen that the curve of Supplier 4 started to climb above the curve of Supplier 5 after the third rank.

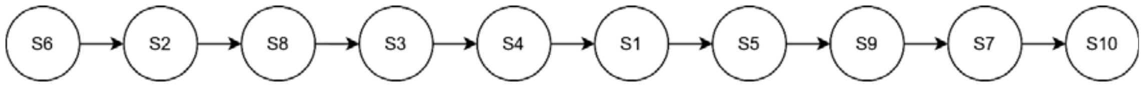


Figure 5.10. Outranking diagram of suppliers with ROC weights of ranks

Weighted scores θ_i calculated by RS, ROC and RR rank weighting approaches, and corresponding supplier rankings are reported in Table 5.26 and Table 5.27. They show the results for DMs who consider all ranks and who consider just the first three ranks, respectively.

Table 5.26. θ_i scores and final rankings of suppliers when all ranks are considered

Rank	Supplier	θ_i (RS)	Supplier	θ_i (ROC)	Supplier	θ_i (RR)
1	S6	0.149	S6	0.201	S6	0.218
2	S2	0.144	S2	0.189	S2	0.201
3	S8	0.134	S8	0.128	S3	0.114
4	S3	0.126	S3	0.127	S8	0.106
5	S1	0.089	S4	0.076	S4	0.073
6	S4	0.088	S1	0.071	S5	0.070
7	S9	0.083	S5	0.068	S1	0.065
8	S5	0.080	S9	0.067	S9	0.063
9	S7	0.064	S7	0.046	S7	0.051
10	S10	0.043	S10	0.027	S10	0.040

For a risk-averse DM who prefers to consider the supplier performances in all ranks, results in Table 5.26 can be used. We can see that Suppliers 6 and 2 are the best two alternatives for all weighting approaches. They are followed by Suppliers 8 and 3, however their positions are reversed in RS/ROC and RR. Since there are larger differences between RR weights of the first ranks, Supplier 3, which has higher probabilities on the first and second ranks than Supplier 8, is placed in the third overall rank with RR weighting approach.

Table 5.27. θ_i scores and final rankings of suppliers when the first three ranks are considered

Rank	Supplier	θ_i (RS)	Supplier	θ_i (ROC)	Supplier	θ_i (RR)
1	S6	0.312	S6	0.352	S6	0.322
2	S2	0.290	S2	0.317	S2	0.295
3	S3	0.135	S3	0.119	S3	0.126
4	S8	0.108	S8	0.083	S8	0.105
5	S5	0.053	S5	0.047	S5	0.053
6	S4	0.045	S4	0.039	S4	0.044
7	S1	0.028	S1	0.022	S1	0.026
8	S9	0.022	S9	0.016	S9	0.022
9	S7	0.007	S7	0.004	S7	0.007
10	S10	0.000	S10	0.000	S10	0.000

On the other hand, for risk-seeking DMs, results of Table 5.27 are more useful to consult since they include rankings when only the first three ranks are considered. For DMs who want to decide based on the performances just for the first three ranks, again Suppliers 6 and 2 are the best choices. Now, for all ranking procedures Supplier 3 is better than Supplier 8 since the performances on lower ranks are ignored. In Table 5.26, when all ranks are important, Supplier 5 is worse than Supplier 1 and 4, but now since the lower ranks are ignored, Supplier 5 is placed on the fifth rank. Additionally, all weighting procedures result in the same ranking list; therefore, we may say that with lower number of ranks, weighting approaches tend to provide similar results.

We also test for the confidence of the supplier ranking obtained by the score-based approach. Firstly, ANOVA results are gathered and reported in Table 5.28 and they show that at least one supplier exists with a net flow mean which is significantly different from other suppliers.

Table 5.28. ANOVA results for the means of net flows of suppliers

Source	Df	Sum Sq.	Mean Sq.	F Value	P Value
Suppliers	9	1366	151.75	88.65	<2e-16
Residuals	990	1695	1.71		

Then, to test the consecutive supplier pairs in the ranking given in Figure 5.10, Tukey's Test is performed. Table 5.29 summarizes the results.

Table 5.29. Tukey's Test results for net flows of supplier pairs

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
S1	-									
S2	(1.59;2.77) 2.18*	-								
S3	(0.57;1.75) 1.16*	(-1.61;-0.43) -1.02*	-							
S4	(-0.63;0.55) -0.04	(-2.81;-1.63) -2.22*	(-1.79;-0.61) -1.20*	-						
S5	(-0.72;0.45) -0.14	(-2.90;-1.73) -2.32*	(-1.88;-0.71) -1.30*	(-0.68;0.49) -0.10	-					
S6	(1.74;2.92) 2.33*	(-0.44;0.74) 0.15	(0.58;1.76) 1.17*	(1.78;2.96) 2.37*	(1.88;3.05) 2.47*	-				
S7	(-1.44;-0.27) -0.85*	(-3.62;-2.45) -3.03*	(-2.60;-1.43) -2.01*	(-1.40;-0.23) -0.81*	(-1.30;-0.13) -0.72*	(-3.77;-2.60) -3.18*	-			
S8	(0.65;1.82) 1.23*	(-1.53;-0.36) -0.95*	(-0.51;0.66) 0.07	(0.68;1.86) 1.27*	(0.78;1.96) 1.37*	(-1.69;-0.51) -1.10*	(1.50;2.67) 2.09*	-		
S9	(-0.88;0.29) -0.30	(-3.06;-1.89) -2.48*	(-2.04;-0.87) -1.46*	(-0.85;0.33) -0.26	(-0.75;0.43) -0.16	(-3.22;-2.04) -2.63*	(-0.03;1.14) 0.56	(-2.12;-0.94) -1.53*	-	
S10	(-1.86;-0.69) -1.27*	(-4.04;-2.87) -3.45*	(-3.02;-1.85) -2.43*	(-1.82;-0.65) -1.24*	(-1.72;-0.55) -1.14*	(-4.19;-3.02) -3.61*	(-1.01;0.17) -0.42	(-3.09;-1.92) -2.51*	(-1.56;-0.39) -0.98*	-

According to Table 5.28, we can make some conclusions. For example, Supplier 2 is ranked higher than Supplier 1 and the results in the table confirm this since the mean difference of Supplier 2 is higher than Supplier 1 by 2.18. Also, the mean differences between Supplier 6, which is the best alternative in Figure 5.10, and other suppliers are positive. However, when we look at the confidence intervals, we can see that some of the consecutive pairs in the ranking are not significantly different from each other such as S1 and S4, S1 and S5, S2 and S6, S3 and S8, and S7 and S9. The best two suppliers, Suppliers 6 and 2 can be considered as similar. Therefore, we can say that these additional analyses are useful to ensure alternatives are robustly discriminated when the score-based approach is performed.

5.2.3. Results of Probabilistic PROMETHEE Approach for Supplier Selection

We implement the three versions of the probabilistic PROMETHEE approach for supplier selection case. Table 5.30 gives outranking values calculated by the first version. The partial ranking by PROMETHEE I rules has some incomparability relationships between suppliers as shown in Figure 5.11.

Table 5.30. Outranking values by the first version of probabilistic PROMETHEE for Supplier Selection

	φ^+	φ^-	φ
S1	2.463	3.109	-0.646
S2	4.150	1.961	2.189
S3	3.338	2.399	0.939
S4	2.987	2.563	0.424
S5	3.051	3.034	0.017
S6	3.806	2.208	1.597
S7	2.496	3.861	-1.364
S8	4.111	2.538	1.573
S9	2.043	3.232	-1.188
S10	1.521	5.063	-3.542

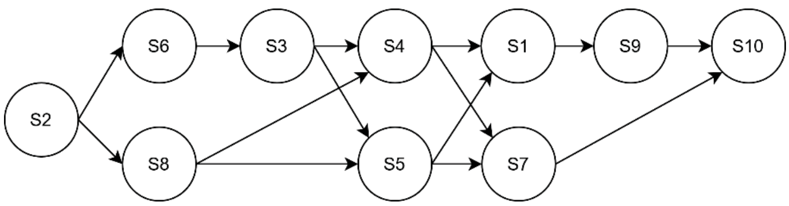


Figure 5.11. Partial ranking of suppliers by the first version of probabilistic PROMETHEE

The best alternative is Supplier 2, followed by Suppliers 6 and 8, whereas the worst alternative is Supplier 10. Since there is not any dominance relation between Suppliers 8 and 6, and Suppliers 8 and 3, we conclude that they are incomparable. Thus, the DM can select Supplier 6 or 3 instead of Supplier 8 according to his/her preferences. Suppliers 4 and 5 have also incomparability relationship whereas Supplier 7 is incomparable with Suppliers 1 and 9.

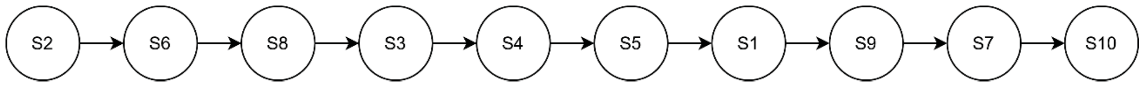


Figure 5.12. Complete ranking of suppliers by the first version of probabilistic PROMETHEE

Figure 5.12 shows the complete ranking of suppliers for the first version. Generally, the best and the worst alternatives are the same, but this time, we have strict ranking for the incomparable suppliers of the partial ranking. Suppliers 2, 6 and 8 are the best three suppliers, whereas Supplier 10 is again the worst one.

For the second version of the problem, threshold value for quality system certifications, flexibility, cooperation and reputation is determined as 2. In Table 5.31, outranking values of the second approach are reported, and Figure 5.13 and Figure 5.14 show the partial and complete ranking of suppliers obtained by the second version of our probabilistic PROMETHEE approach, respectively.

Table 5.31. Outranking values by the second version of probabilistic PROMETHEE for Supplier Selection

	φ^+	φ^-	φ
S1	2.266	2.597	-0.331
S2	3.593	1.847	1.746
S3	2.932	2.123	0.809
S4	2.525	2.187	0.338
S5	2.720	2.517	0.203
S6	3.702	1.781	1.922
S7	2.265	3.566	-1.301
S8	3.524	2.414	1.111
S9	1.753	2.878	-1.125
S10	1.254	4.624	-3.370

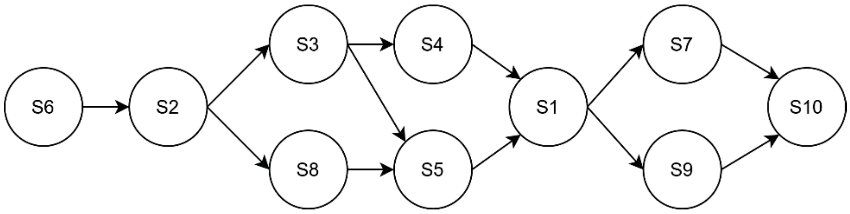


Figure 5.13. Partial ranking of suppliers by the second version of probabilistic PROMETHEE

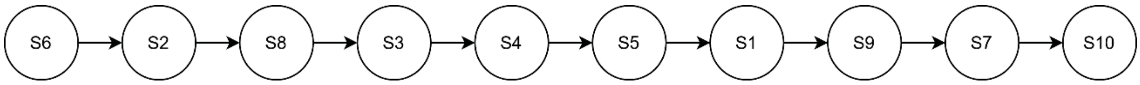


Figure 5.14. Complete ranking of suppliers by the second version of probabilistic PROMETHEE

For both partial and complete rankings, Supplier 6 is the best alternative, followed by Supplier 2, and the worst one is again Supplier 10. Incomparability relationships between Suppliers 8 and 3, 8 and 4, 4 and 5, and 7 and 9 in the partial ranking are solved in the complete ranking for the DMs who prefer strict ranking. When we compare the results of the second approach with the first approach, we conclude that Supplier 2 is better than Supplier 6 with small differences in the first version of the approach, since in the second version, Supplier 6 dominates Supplier 2 when we define a threshold value of 2 for some criteria. Also, we deduce that Supplier 6 dominates Supplier 8 with at least 2 level differences in the defined criteria since Suppliers 6 and 8 are incomparable in the partial ranking of the first approach, whereas Supplier 6 dominates Supplier 8 in the second version of the approach.

Table 5.32 reports the outranking values of suppliers by the third version of the approach. Figure 5.15 and Figure 5.16 illustrate the partial and complete ranking of suppliers, respectively.

Table 5.32. Outranking values by the third version of probabilistic PROMETHEE for Supplier Selection

	φ^+	φ^-	φ
S1	0.286	0.390	-0.103
S2	0.533	0.255	0.278
S3	0.433	0.292	0.141
S4	0.374	0.314	0.060
S5	0.381	0.371	0.010
S6	0.496	0.302	0.194
S7	0.311	0.640	-0.329
S8	0.675	0.301	0.375
S9	0.244	0.406	-0.162
S10	0.220	0.683	-0.463

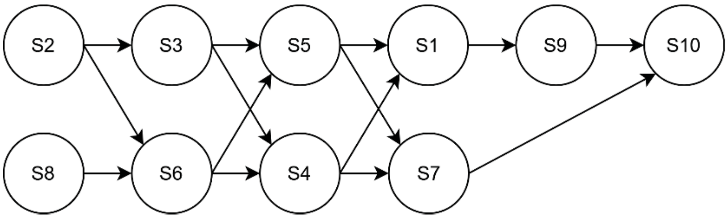


Figure 5.15. Partial ranking of suppliers by the third version of probabilistic PROMETHEE

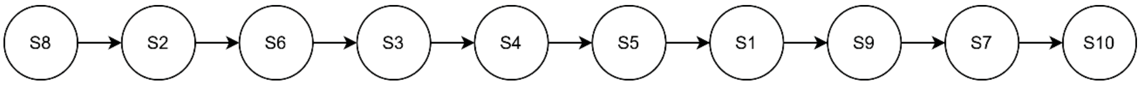


Figure 5.16. Complete ranking of suppliers by the third version of probabilistic PROMETHEE

This time, Suppliers 2 and 8 are the best alternatives and they are incomparable with each other in the partial ranking. Also, Supplier 8 has an incomparability relationship with Supplier 3, whereas supplier pairs 3 and 6, 5 and 4, 1 and 7 and, 7 and 9 are found as incomparable since they cannot dominate each other based on the positive and negative outranking values. However, the complete ranking presents a strict ranking for these incomparable suppliers. For both partial and complete rankings, the best and worst suppliers are the same. If we compare the results of the third approach with the previous ones, we conclude that Supplier 8 performs better than Suppliers 2 and 6 with high differences, since the third approach gives more importance to higher levels of differences.

5.2.4. Sensitivity Analysis on Criteria Weights for Supplier Selection

In this case study, the original weights of criteria are derived by AHP. First, we study whether relatively small changes in weights result in big differences in results. Rather than changing the AHP weights randomly, we apply ROC weighting with the same supplier selection expert. Table 5.33 presents the resulting ROC weights. The most evident difference is in the weight of cost, which is now quite lower. There are differences in the weights of other criteria as well.

Table 5.33.Weight of criteria by ROC for Supplier Selection

Criteria	Product Quality	Cost	Delivery Perf.	Quality Syst. Cert.	Flexibility	Cooperation	Reputation
Weights	0.370	0.156	0.228	0.109	0.073	0.044	0.020

Figure 5.17 illustrates the outranking relations for the test-based approach calculated with the new weights. S2 and S6 are still the best suppliers that cannot outrank each other, and S10 is still the worst supplier. S3, which cannot outrank S8 with AHP weights, now outranks it with ROC weights, and it cannot be outranked with S2. There are some other differences, but the general picture is not substantially different from Figure 5.8, and there are no reversals in outranking relations.

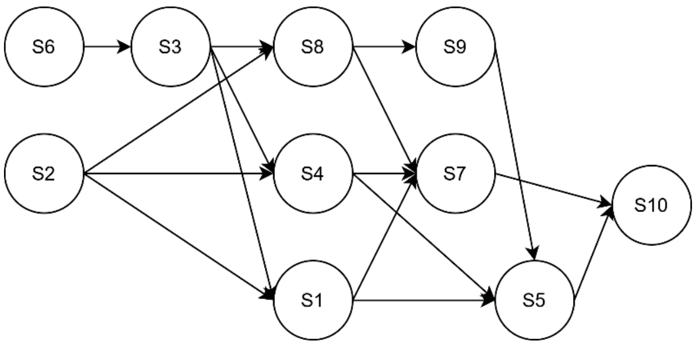


Figure 5.17. Test-based outranking relation of suppliers with ROC weights

The new ranking results of the score-based approach when all ranks are considered are given in Table 5.34. To test the similarity of the rankings with AHP and ROC weights, we use Kendall rank correlation coefficient (Kendall’s Tau). Kendall’s Tau is a measure to assess the linear correlation between two ranks. It returns a value between -1 and 1. Values closer to -1 imply that rankings have negative correlation and values closer to 1 show that rankings have positive correlation, thus, they are similar. A value of 0 means that your rankings are not correlated. Kendall's Tau between the old and new rankings of RS, ROC and RR are 0.822, 0.778 and 0.822, respectively. We can conclude that the rankings stay reasonably similar when the weights are changed.

Table 5.34. θ_i measures and final rankings by ROC when all ranks are considered

Rank	Supplier	θ_i (RS)	Supplier	θ_i (ROC)	Supplier	θ_i (RR)
1	S6	0.143	S2	0.185	S2	0.196
2	S3	0.142	S6	0.184	S6	0.194
3	S2	0.142	S3	0.161	S3	0.153
4	S8	0.109	S4	0.093	S4	0.085
5	S4	0.101	S8	0.089	S8	0.074
6	S1	0.095	S1	0.079	S1	0.073
7	S9	0.087	S9	0.072	S9	0.067
8	S7	0.076	S7	0.058	S5	0.062
9	S5	0.069	S5	0.057	S7	0.058
10	S10	0.038	S10	0.023	S10	0.039

When only the first three ranks are considered, the ranking of the suppliers for all weighing methods is S2–S6–S3–S4–S1–S5–S8–S9–S7–S10. Kendall's Tau between this ranking and the ranking obtained with AHP weights in Table 5.27 is 0.733.

We also obtain results of the probabilistic PROMETHEE approach with the new ROC weights. Figure 5.18 illustrates the new partial ranking of suppliers by the first version of the probabilistic PROMETHEE. When we compare it with the ranking in Figure 5.11, we conclude that the ranking has not substantially changed. Supplier 2 is still the best and Suppler 10 is still the worst alternative. With the new weights, Supplier 3 outranks Supplier 6. Also, supplier 4 outranks Supplier 5, whereas they are incomparable in Figure 5.11.

On the other hand, the new complete ranking for the first version is found as S2–S3–S6–S8–S4–S1–S5–S9–S7–S10, and Kendall's Tau between this ranking and the ranking in Figure 5.12 is 0.867. Thus, we conclude that the complete rankings of the first version by AHP and ROC weights are reasonably similar.

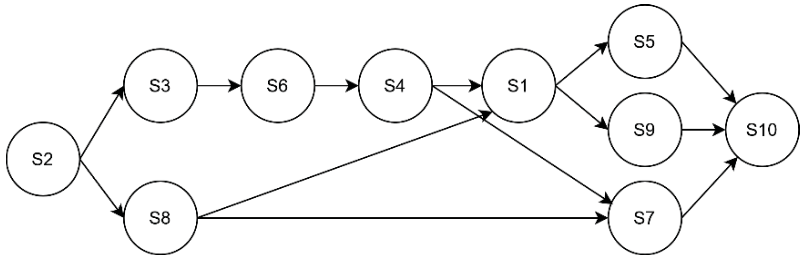


Figure 5.18. Partial ranking of suppliers by the first version of the probabilistic PROMETHEE with ROC weights

Figure 5.19 shows the new partial ranking of the second version of the approach. This time, Suppliers 2 and 6, and 5 and 7 are incomparable, whereas Supplier 6 outranks Supplier 2 and Supplier 5 outranks Supplier 7 in the partial ranking shown in Figure 5.13. Again, there are some differences for suppliers with incomparability relationships, but the ranking stays reasonable similar. To assess the similarity of two partial ranking, Kendall's Tau is used in some cases by assigning same ranks for indifferent alternatives. However, in our case it is not possible to follow that approach since our partial ranking includes incomparable relations. We cannot assign same ranks to incomparable alternatives because it leads to wrong conclusions. For example, in the ranking in Figure 5.19, S6 is incomparable with both S2 and S3 but, S2 outranks S3. In that situation, we cannot assign same rank for those three suppliers.

The complete ranking of the second version calculated with ROC weights is S2–S6–S3–S4–S8–S1–S5–S9–S7–S10. This ranking and the ranking in Figure 5.14 are similar since the Kendall's Tau between them is found as 0.822.

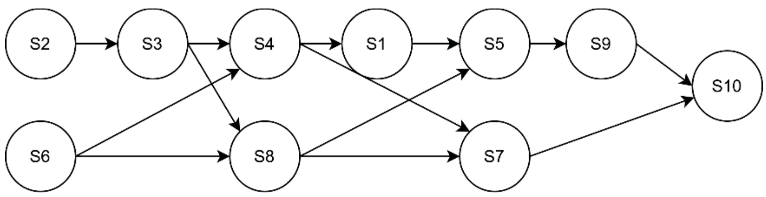


Figure 5.19.Partial ranking of suppliers by the second version of the probabilistic PROMETHEE with ROC weights

In Figure 5.20, the new partial ranking of the third version is illustrated. It has small differences from the ranking in Figure 5.15. Suppliers 2 and 8 are still the best alternatives that cannot outrank each other. This time, Suppliers 10 and 7 are incomparable and they are the worst suppliers. The other difference is between Suppliers 4 and 5. In the ranking by AHP weights they are incomparable, but with ROC weights Supplier 4 outranks Supplier 5.

S2–S8–S3–S6–S4–S1–S5–S9–S7–S10 is the new complete ranking obtained using ROC weights. Generally, the new ranking is similar to the one in Figure 5.16. As in the previous version, the first and the second suppliers and, the third and the fourth suppliers change places with each other. Kendall's Tau between these two rankings is 0.867.

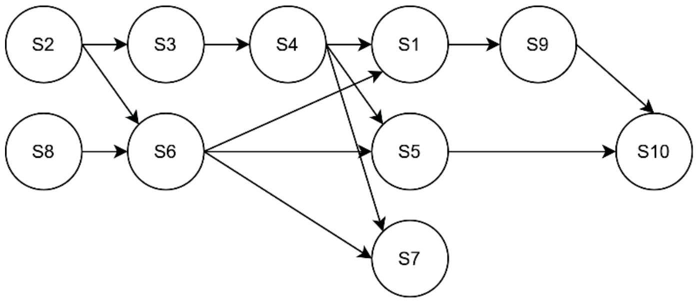


Figure 5.20. Partial ranking of suppliers by the third version of the probabilistic PROMETHEE with ROC weights

As in the shoulder pain case, we also conduct weight stability interval analysis for the supplier selection with the score-based approach. We follow the same approach using the rules introduced by Mareschal (1988). Table 5.35 shows the weight intervals of each criterion which will not change the supplier rankings and the scores of score-based approach. Since product quality, cost and delivery performance criteria have relatively narrow ranges, their weights should be determined carefully. However, the ranking is not so sensitive to the weights of other criteria. Thus, uncertainties in these criteria can be tolerated better.

Table 5.35. Weight stability intervals for AHP weights

Criteria	AHP Weight	Weight Stability Interval
Product Quality	0.356	[0.337 – 0.360]
Cost	0.231	[0.216 – 0.264]
Delivery Performance	0.230	[0.215 – 0.279]
Quality System Certificates	0.075	[0.015 – 0.162]
Flexibility	0.032	[0.000 – 0.392]
Cooperation	0.042	[0.000 – 0.303]
Reputation	0.034	[0.000 – 0.216]

6. CONCLUSIONS

In this thesis, we proposed test-based, score-based and probabilistic PROMETHEE approaches for MCDM under uncertainty. The proposed approaches use uncertain data in criteria evaluations which may be elicited from previous observations, samples from probabilistic models or judgements of experts. The approaches provide different levels of precision and flexibility to evaluate solutions. Our test-based approach modifies PROMETHEE I to work with uncertain data and it assesses error rates in outranking scores. It is useful for DMs who want to rank alternative solutions according to their performance without forcing a strict ranking. Our score-based approach is based on PROMETHEE II, and it uses probabilities of solutions to occupy each possible rank. DMs can observe a summary of performances of solutions with cumulative distributions. Also, it provides weighted scores to rank solutions completely for both risk-seeking and risk-averse DMs. Our third approach, the probabilistic PROMETHEE uses joint probabilities instead of sampling from distributions. It focuses on probabilities where an alternative has better criterion values than the other one. It has different versions that provide the DM the opportunity to differentiate between different levels of differences in criteria values. Our approaches present systematic ways to work with probabilistic criteria, elicit the preferences of the DM regarding the importance of criteria and obtain overall assessments of solutions in an uncertain environment.

We tested our approaches using two different case studies: a treatment selection problem for shoulder pain and a supplier selection problem. For the case of shoulder pain, we used probability distributions of treatment alternatives in each criterion produced from the information obtained from physiotherapists. But we suggest using posteriors of a BN model to have more accurate results. For the case of supplier selection, we used posterior probabilities of a BN model developed for that case by domain experts. Using posterior probabilities generated by a BN model in our approaches is advantageous since specific case-based results are obtained. Our studies illustrate how BN outputs can be processed by MCDM methods to provide decision support. In addition, we conducted sensitivity analysis on the weights of criteria; this analysis can be used to determine which criteria need the most careful evaluation.

In future studies, different types of DM behaviour, which would result in different preference functions and parameters, can be explored. Indirect elicitation approaches for weights can be implemented. Applications with patients can be conducted to test the approaches in actual medical problems.

7. REFERENCES

Adunlin, G., Diaby, V., Montero, A.J., Xiao, H., Multicriteria decision analysis in oncology, Heal. Expect., 18 (**2015**) 1812–1826.

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

Baležentis, T., Streimikiene, D., Multi-criteria ranking of energy generation scenarios with Monte Carlo simulation, Appl. Energy, 185 (**2017**) 862–871.

Baudry, G., Macharis, C., Vallée, T., Range-based Multi-Actor Multi-Criteria Analysis: A combined method of Multi-Actor Multi-Criteria Analysis and Monte Carlo simulation to support participatory decision making under uncertainty, Eur. J. Oper. Res., 264 (**2018**) 257–269.

Benetto, E., Dujet, C., Rousseaux, P., Integrating fuzzy multicriteria analysis and uncertainty evaluation in life cycle assessment, Environ. Model. Softw., 23 (**2008**) 1461–1467.

Betrie, G.D., Sadiq, R., Morin, K.A., Tesfamariam, S., Selection of remedial alternatives for mine sites : A multicriteria decision analysis approach, J. Environ. Management, 119 (**2013**) 36–46.

Beynon, M.J., Wells, P., The lean improvement of the chemical emissions of motor vehicles based on preference ranking: A PROMETHEE uncertainty analysis, Omega, 36 (**2008**) 384–394.

Brans, J.P., Vincke, P., Mareschal, B., How to select and how to rank projects : The PROMETHEE method, Eur. J. Oper. Res., 24 (**1986**) 228–238.

Broekhuizen, H., G.M., C., Oudshoorn, G., A Review and Classification of Approaches for Dealing with Uncertainty in Multi-Criteria Decision Analysis for Healthcare Decisions, Pharmacoeconomics, 33 (**2015**) 445–455.

Cavalcante, C.A.V., De Almeida, A.T., A multi-criteria decision-aiding model using PROMETHEE III for preventive maintenance planning under uncertain conditions, J. Qual. Maint. Eng., 13 (**2007**) 385–397.

Chai, J., Liu, J.N.K., Ngai, E.W.T., Application of decision-making techniques in supplier selection: A systematic review of literature, Expert Syst. Appl., 40 (**2013**) 3872–3885.

D'Avignon, G.R., Vincke, P., An outranking method under uncertainty, Eur. J. Oper. Res., 36 (**1988**) 311–321.

Diaby, V., Goeree, R., How to use multi-criteria decision analysis methods for reimbursement decision-making in healthcare: A step-by-step guide, Expert Rev. Pharmacoeconomics Outcomes Res., 14 (**2014**) 81–99.

Diaby, V., Sanogo, V., Moussa, K.R., ELICIT: An alternative imprecise weight elicitation technique for use in multi-criteria decision analysis for healthcare, Expert Rev. Pharmacoeconomics Outcomes Res., 16 (**2016**) 141–147.

Dolan, J.G., Veazie, P.J., The Feasibility of Sophisticated Multicriteria Support for Clinical Decisions, Med. Decis. Making, 38 (**2018**) 465–475.

Dong, B., Gao, P., Wang, H., Liao, S., Clustering Human Wrist Pulse Signals via Multiple Criteria Decision Making, Int. Conf. on Tools with Artif. Intell (ICTAI), December, 243–250, **2014**.

Dorini, G., Kapelan, Z., Azapagic, A., Managing uncertainty in multiple-criteria decision making related to sustainability assessment, Clean Technol. Environ. Policy, 13 (**2011**) 133–139.

Doumpos, M., Zopounidis, C., A multicriteria decision support system for bank rating, Decis. Support Syst., 50 (**2010**) 55–63.

Durbach, I.N., Outranking under uncertainty using scenarios, Eur. J. Oper. Res., 232 (**2014**) 98–108.

Durbach, I.N., Stewart, T.J., Modeling uncertainty in multi-criteria decision analysis, Eur. J. Oper. Res., 223 (**2012**) 1–14.

Fenton, N., Neil, M., Risk assessment and decision analysis with bayesian networks, CRC Press, **2012**.

Gervásio, H., Simões Da Silva, L., A probabilistic decision-making approach for the sustainable assessment of infrastructures, Expert Syst. Appl., 39 (**2012**) 7121–7131.

Ghosh, B., Assessing surgical performance in healthcare institutions using fuzzy composite programming, IEEE Reg. 10 Colloq. and 3rd Int. Conf. on Ind. and Inf. Syst. (ICIIS), 8-10 December, Kharagpur, India, 1–6, **2008**.

Govindan, K., Rajendran, S., Sarkis, J., Murugesan, P., Multi criteria decision making approaches for green supplier evaluation and selection: A literature review, J. Clean.

Prod., 98 (**2015**) 66–83.

Hopfe, C.J., Augenbroe, G.L.M., Hensen, J.L.M., Multi-criteria decision making under uncertainty in building performance assessment, Build. Environ., 69 (**2013**) 81–90.

Hyde, K., Maier, H.R., Colby, C., Incorporating uncertainty in the PROMETHEE MCDA method, J. Multi-Criteria Decis. Anal., 12 (**2003**) 245–259.

Hyde, K.M., Maier, H.R., Colby, C.B., A distance-based uncertainty analysis approach to multi-criteria decision analysis for water resource decision making, J. Environ. Management, 77 (**2005**) 278–290.

Jacquet-Lagreze, E., An application of the UTA Discriminant Model for the Evaluation of R&D Project, Advances in Multicriteria Analysis, Pardalos P. M., Siskos Y., C.Z. (Eds.), Kluwer Academic Publishers, **1985**.

Jacquet-Lagreze, E., Siskos, J., Assessing a set of additive utility functions for multicriteria decision-making, the UTA method, Eur. J. Oper. Res. 10 (**1982**) 151–164.

Kaya, R., Yet, B., Building Bayesian networks based on DEMATEL for multiple criteria decision problems: A supplier selection case study, Expert Syst. Appl., 134 (**2019**) 234–248.

Keršuliene, V., Zavadskas, E.K., Turskis, Z., Selection of rational dispute resolution method by applying new step - wise weight assessment ratio analysis (Swara) By Applying New Step-Wise Weight Assessment Ratio, J. Bus. Econ. Management, 11 (**2011**) 243-258.

Kilic, H.S., Zaim, S., Delen, D., Development of a hybrid methodology for ERP system selection: The case of Turkish Airlines, Decis. Support Syst., 66 (**2014**) 82–92.

Kuang, H., Hipel, K.W., Kilgour, D.M., Evaluation of Source Water Protection Strategies in Waterloo Region based on Grey Systems Theory and PROMETHEE II, in: Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics, **2012**, pp. 2775–2779.

Kulak, O., Goren, H.G., Supciller, A.A., A new multi criteria decision making approach for medical imaging systems considering risk factors, Appl. Soft Comput. Journal, 35 (**2015**) 931–941.

Kuo, R.J., Wu, Y.H., Hsu, T.S., Integration of fuzzy set theory and TOPSIS into HFMEA to improve outpatient service for elderly patients in Taiwan, J. Chinese Med. Assoc., 75 (**2012**) 341–348.

La Scalia, G., Aiello, G., Rastellini, C., Micale, R., Cicalese, L., Multi-Criteria Decision Making support system for pancreatic islet transplantation, Expert Syst. Appl., 38 (**2011**) 3091–3097.

Lauritzen, S.L., Spiegelhalter, D.J., Local Computations with Probabilities on Graphical Structures and Their Application to Expert Systems, Journal of Royal Stat. Soc. Ser. B, 50 (**1988**) 157–194.

Levary, R.R., Wan, K., A simulation approach for handling uncertainty in the analytic hierarchy process, Eur. J. Oper. Res., 106 (**1998**) 116–122.

Li, D.P., He, J.Q., Cheng, P.F., Wang, J.Q., Zhang, H.Y., A novel selection model of surgical treatments for early gastric cancer patients based on heterogeneous multicriteria group decision-making, Symmetry (Basel)., 10 (**2018**).

Liu, Y., Fan, Z.P., Zhang, Y., A method for stochastic multiple criteria decision making based on dominance degrees, Inf. Sci. (Ny), 181 (**2011**) 4139–4153.

Lupo, T., A fuzzy framework to evaluate service quality in the healthcare industry: An empirical case of public hospital service evaluation in Sicily, Appl. Soft Comput. Journal, 40 (**2016**) 468–478.

Mareschal, B., Weight stability intervals in multicriteria decision aid, Eur. J. Oper. Res., 33 (**1988**) 54–64.

Mareschal, B., Stochastic multicriteria decision making and uncertainty, Eur. J. Oper. Res., 26 (**1986**) 58–64.

Momani, A.M., Ahmed, A.A., Material Handling Equipment Selection using Hybrid Monte Carlo Simulation and Analytic Hierarchy Process, Int. J. Ind. Manuf., 5 (**2011**) 2177–2182.

Montazar, A., Gheidari, O.N., Snyder, R.L., A fuzzy analytical hierarchy methodology for the performance assessment of irrigation projects, Agric. Water Manag, 121 (**2013**) 113–123.

Montazer, G.A., Saremi, H.Q., Ramezani, M., Design a new mixed expert decision aiding system using fuzzy ELECTRE III method for vendor selection, Expert Syst. Appl., 36 (**2009**) 10837–10847.

Nilashi, M., Ahmadi, H., Ahani, A., Ravangard, R., Ibrahim, O. bin,. Determining the importance of Hospital Information System adoption factors using Fuzzy Analytic Network Process (ANP), Technol. Forecast. Soc. Change, 111 (**2016**) 244–264.

Nobre, F.F., Trotta, L.T.F., Gomes, L.F.A.M., Multi-criteria decision making - An approach to setting priorities in health care, Stat. Med., 18 (**1999**) 3345–3354.

Özkan, A., Evaluation of healthcare waste treatment/disposal alternatives by using multi-criteria decision-making techniques, Waste Manag. Res., 31 (**2013**) 141–149.

Pitchipoo, P., Venkumar, P., Rajakarunakaran, S., Fuzzy hybrid decision model for supplier evaluation and selection, Int. J. Prod. Res., 51 (**2013**) 3903–3919.

Pomerol, J.-C., Barba-Romero, S., Multicriterion Decision in Management Principles and Practice, Springer Science & Business Media, Springer US, Boston, MA, **2000**.

Rahimi, S., Gandy, L., Mogharreban, N., A web-based high-performance multicriteria decision support system for medical diagnosis, Int. J. Intell. Syst., 22 (**2007**)1083–1099.

Rietveld, P., Ouwersloot, H., Ordinal data in multicriteria decision making, a stochastic dominance approach to siting nuclear power plants, Eur. J. Oper. Res., 56 (**1992**) 249–262.

Roszkowska, E., Rank Ordering Criteria Weighting Methods – A Comparative Overview, Optim. Stud. Ekon., 5 (**2013**) 14–33.

Roy, B., Paradigm and Challanges, in: Figueira, J., Greco, S., Ehrgott, M. (Eds.), Multiple Criteria Decision Analysis State of the Art Surveys. Springer US, Boston, **2005**, pp. 446–470.

Saaty, T.L., How to make a decision: The Analytic Hierarchy Process, Eur. J. Oper. Res., 48 (**1990**) 9–26.

Sałabun, W., Piegat, A., Comparative analysis of MCDM methods for the assessment of mortality in patients with acute coronary syndrome, Artif. Intell. Rev., 48 (**2017**) 557–571.

Siskos, J., Yannacopoulos, D., UTASTAR: An ordinal regression method for building additive value functions, Investigaçao Operacional, 5 (**1985**) 39–53.

Souissi, S. Ben, Abed, M., Elhiki, L., Fortemps, P., Pirlot, M., Categorizing the suitability of an alternative for a subject, IEEE Symp. Ser. Comput. Intell, **2016**.

Stewart, T.J., Dealing with Uncertainties in MCDA, in: Figueira, J., Greco, S., Ehrgott, M. (Eds.), Multiple Criteria Decision Analysis State of the Art Surveys. Springer US, Boston, **2005**, pp. 446–470.

Sureeyatanapas, P., Comparison of rank-based weighting methods for multi-criteria decision making, KKU Eng. Journal, 43 (**2016**) 376–379.

Taghipour, S., Banjevic, D., Jardine, A.K.S., Prioritization of medical equipment for maintenance decisions, J. Oper. Res. Soc, 62 (**2011**) 1666–1687.

Tuncer Şakar, C., Yet, B., Çok Kriterli Karar Verme Problemlerinde Fayda Fonksiyonu Ağirliklarinin Tahmin Edilmesi İçin Matematiksel Model Temelli Bir Yöntem, Uludağ Univ. J. Fac. Eng., 23 (**2018**) 379–402.

Von Winterfeldt, D. and Edwards, W., Decision Analysis and Behavioural Research, Cambridge University Press, Cambridge, UK, **1986**.

Yuen, K.K.F., Ting, T.O., Textbook Selection Using Fuzzy PROMETHEE II Method, Int. J. Futur. Comput. Commun., 1 (**2012**) 76–78.

Zhang, C., Wu, X., Wu, D., Liao, H., Luo, L., Herrera-Viedma, E., An intuitionistic multiplicative ORESTE method for patients' prioritization of hospitalization, Int. J. Environ. Res. Public Health, 15 (**2018**) 1–18.

Zhang, Y., Fan, Z.P., Liu, Y., A method based on stochastic dominance degrees for stochastic multiple criteria decision making, Comput. Ind. Eng., 58 (**2010**) 544–552.

Zimmer, K., Fröhling, M., Schultmann, F., Sustainable supplier management - a review of models supporting sustainable supplier selection, monitoring and development, Int. J. Prod. Res., 54 (**2016**) 1412–1442.