



university of  
 groningen



---

Master Thesis

# Pairwise- and mixed treatment comparison models in multicriteria benefit-risk analysis

Author:

**Jing Zhao**

[jingzhao@kth.se](mailto:jingzhao@kth.se)

Supervisors :

Henrik Boström, Department of Computer and Systems Sciences, Stockholm University  
Tommi Tervonen, Faculty of Economics and Business, University of Groningen

Engineering and Management of Information Systems  
Royal Institute of Technology  
In collaboration with Department of Computer and Systems Sciences  
Stockholm University

Stockholm  
June, 2010

---

This thesis corresponds to 20 weeks of full-time work for the author.

## Abstract

Benefit-Risk (BR) analysis plays an important role in evidence based medicine. BR analysis is done to evaluate the safety and efficacy of different medical compounds in, for example, drug regulation and drug development processes. In order to generate benefit and risk profiles, evidence synthesis is necessary to estimate effects from available evidence. Currently evidence synthesis is mostly done through meta-analysis that allows only comparisons between two drugs on a single criterion. The recently proposed mixed treatment comparison (MTC) method allows synthesizing all the available evidence through application of a Bayesian evidence network. Decision making problems in the context of drug BR analysis often involve multiple criteria and uncertainties, therefore multicriteria decision analysis (MCDA) methods are needed to identify the trade-offs among different medical compounds. Stochastic Multicriteria Acceptability Analysis (SMAA), a family of MCDA methods, has previously been applied with criteria measurements from meta-analysis to analyze the relative BR profiles of four second-generation antidepressants. In this thesis, a new SMAA model with criteria measurements from MTC is proposed and evaluated with the same antidepressants. The MTC based SMAA model is compared with the meta-analysis based SMAA model quantitatively and qualitatively. This thesis concludes that the MTC based SMAA model gives more discriminative results than the meta-analysis based SMAA model does, and improves the transparency of decision making process.

**Key words:** meta-analysis; mixed treatment comparison (MTC); stochastic multicriteria acceptability analysis (SMAA); drug benefit-risk analysis; multicriteria decision analysis (MCDA)

## **Acknowledgment**

This master thesis was processed in two universities, Royal Institute of Technology (KTH) in collaboration with Department of Computer and Systems Sciences at Stockholm University in Sweden, and University of Groningen (RuG) in Netherlands. First of all, I would like to thank my supervisors, Prof. Henrik Boström at Stockholm University and Dr. Tommi Tervonen at RuG, for their guidance, support and encouragement from the beginning till the end.

Secondly, I would like to thank all the people I worked with during this thesis process, Gert van Valkenhoef, Prof. Hans Hillege, Prof. Bert de Brock etc., for their valuable suggestions and warm help.

Finally, I want to thank my parents and friends for giving me endless love and concern.

Without their help and support I would never have completed my master thesis. Thank you all from my heart!

## **Glossary**

ADR: Adverse Drug Reaction

BR: Benefit-Risk

CI: Confidence Interval

DM: Decision Maker

HAM-D: Hamilton Rating Scale for Depression

MADRS: Montgomery-Asberg Depression Rating Scale

MAUT: Multi-Attribute Utility Theory

MCDA: Multi-Criteria Decision Analysis

MTC: Mixed Treatment Comparison

NNT: Number Needed to Treat

OR: Odds Ratio

RR: Risk Ratio

SD: Standard Deviation

SMAA: Stochastic Multicriteria Acceptability Analysis

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Background . . . . .	1
1.2	Problem statement . . . . .	4
1.3	Disposition . . . . .	4
<b>2</b>	<b>Extended background</b>	<b>6</b>
2.1	Network meta-analysis . . . . .	6
2.1.1	Consistency model . . . . .	8
2.1.2	Inconsistency model . . . . .	8
2.1.3	Implementation of MTC . . . . .	9
2.2	Stochastic Multicriteria Acceptability Analysis . . . . .	10
2.2.1	SMAA-2 . . . . .	11
2.2.2	Implementation of SMAA . . . . .	16
<b>3</b>	<b>Research methodology</b>	<b>17</b>
3.1	Research type . . . . .	17
3.2	Research steps . . . . .	19
<b>4</b>	<b>Case study</b>	<b>23</b>
4.1	Previous related studies . . . . .	24
4.1.1	Meta-analysis . . . . .	24
4.1.2	MA/SMAA . . . . .	25
4.2	MTC/SMAA . . . . .	27
4.2.1	MTC model . . . . .	28
4.2.2	SMAA model . . . . .	30

4.3	Analysis . . . . .	31
4.3.1	Quantitative comparison . . . . .	31
4.3.2	Qualitative comparison . . . . .	35
<b>5</b>	<b>Concluding remarks</b>	<b>39</b>
5.1	Conclusion . . . . .	39
5.2	Future work . . . . .	40
	<b>References</b>	<b>41</b>
	<b>Appendices</b>	<b>46</b>
	Appendix A . . . . .	46
	Appendix B . . . . .	50
	Appendix C . . . . .	54
	Appendix D . . . . .	55

## List of Figures

1	Network evidence structures. . . . .	7
2	The feasible weight space in the three-criteria case. . . . .	12
3	The feasible weight with constraints on $w_1$ in the three-criteria case. . . . .	15
4	The feasible weight space with complete ranking of the weights in the three-criteria case. . . . .	15
5	Process diagram of case study. . . . .	23
6	Evidence network of existing studies comparing efficacy of second-generation antidepressants. . . . .	24
7	Evidence network of chosen antidepressants. . . . .	25
8	Meta-analysis of Fluoxetine compared with Paroxetine. . . . .	26
9	Diagram of MTC model application. . . . .	29
10	An imaginary example of a study network. . . . .	30
11	Rank acceptability indices without preference information. Left one is from MA/SMAA and right one from MTC/SMAA. . . . .	33
12	Rank acceptability indices in mild depression scenario. Left one is from MA/SMAA and right one from MTC/SMAA. . . . .	33
13	Rank acceptability indices in severe depression scenario. Left one is from MA/SMAA and right one from MTC/SMAA. . . . .	33
14	MA/SMAA, rank acceptability indices for the model without preference information. . . . .	47
15	MA/SMAA, rank acceptability indices from the scenario of mild depression. . . . .	48
16	MA/SMAA, rank acceptability indices from the scenario of severe depression. . . . .	49
17	MTC/SMAA, rank acceptability indices for the model without preference information. . . . .	51

18	MTC/SMAA, rank acceptability indices from the scenario of mild depression. . . . .	52
19	MTC/SMAA, rank acceptability indices from the scenario of severe depression. . . . .	53
20	MTC/SMAA in absolute measurement, rank acceptability indices for the model without preference information. . . . .	55
21	MTC/SMAA in absolute measurement, rank acceptability indices from the scenario of mild depression. . . . .	57
22	MTC/SMAA in absolute measurement, rank acceptability indices from the scenario of severe depression. . . . .	58



## List of Tables

1	Description of related previous studies and this thesis . . . . .	4
2	Consistency model results . . . . .	30
3	Variation of rank acceptability indice in MTC/SMAA compared with MA/SMAA and the value is expressed as rank (difference) . . . . .	34
4	MA/SMAA, criteria characteristics . . . . .	46
5	MA/SMAA, criteria measurements. The values are given as mean $\pm$ standard deviation. The measurement units are as presented in Table 4. . . . .	46
6	MA/SMAA, rank acceptability indices from the analysis without preference information . . . . .	46
7	MA/SMAA, central weights and corresponding confidence factors from the analysis without preference information . . . . .	47
8	MA/SMAA, rank acceptability indices from the scenario of mild depression . . . . .	47
9	MA/SMAA, rank acceptability indices from the scenario of severe depression . . . . .	48
10	MTC/SMAA, criteria characteristics . . . . .	50
11	MTC/SMAA, criteria measurements. The values are given as mean $\pm$ standard deviation. The measurement units are as presented in Table 10. . . . .	50
12	MTC/SMAA, rank acceptability indices from the analysis without preference information . . . . .	50
13	MTC/SMAA, central weight and corresponding confidence factors from the analysis without preference information . . . . .	51
14	MTC/SMAA, rank acceptability indices from the scenario of mild depression . . . . .	51

15	MTC/SMAA, rank acceptability indices from the scenario of severe depression . . . . .	52
16	MTC/SMAA in absolute measurement, criteria characteristics	54
17	MTC/SMAA in absolute measurement, criteria measurements. The values are given as mean $\pm$ standard deviation. The measurement units are as presented in Table 16. . . . .	54
18	MTC/SMAA in absolute measurement, rank acceptability indices from the analysis without preference information . . . .	54
19	MTC/SMAA in absolute measurement, central weight and corresponding confidence factors from the analysis without preference information . . . . .	55
20	MTC/SMAA in absolute measurement, rank acceptability indices from the scenario of mild depression . . . . .	56
21	MTC/SMAA in absolute measurements, rank acceptability indices from the scenario of severe depression . . . . .	56

# 1 Introduction

## 1.1 Background

Drug regulation is aimed at ensuring the safety, quality and efficacy of drugs on the market. Current decision making in drug regulation relies on the expert judgment of the assessors, and such reliance on subjective assessment hides the reasoning supporting the decisions and consequently causes the regulatory process to be insufficiently transparent and traceable.

Nowadays, the regulatory authorities increasingly request an explicit Benefit-Risk (BR) analysis of drugs as it can provide a basis for rational decisions in the use of a particular therapy [1]. Such authorities include the U.S. Food and Drug Administration and the European Medicines Agency. Drug BR analysis can be used for identifying the trade-offs between benefit and risk [2], where benefit is the efficacy of a drug and risk relates to its safety. Drug BR analysis is done everyday by health care professionals, such as regulators, practicing physicians, and employees of insurance companies, to evaluate the safety and efficacy of different medical compounds [3]. If there is only one measure of efficacy and one measure of safety, the BR analysis can be assisted “easily” by plotting possible measurements on a plane [4]. However, normally multiple (more than two) risk criteria have to be taken into account and the standard two-dimensional visualization techniques cannot be applied.

In order to assess the strength of evidence of the benefits and risks of treatments, evidence based medicine applies the available evidence to medical decision making [5]. Evidence synthesis, which became the main issue in evidence based medicine, is a set of formal processes for bringing together different types of evidences so that we can be clear about what we know from research and how we know it [6].

Meta-analysis is a statistical method for quantitatively synthesizing evidence from multiple trials in order to obtain overall pooled effect estimates [7, 8, 9]. It is based on pairwise treatment comparisons. Often the trials compare different treatments in a direct way, but as the number of available treatments

increases the number of possible pairwise comparisons increases quadratically [7]. The pairwise comparison may be inapplicable due to it disallowing indirect comparisons, and consequently it is impossible to decide upon the “best” treatment in a class with no common comparator. For instance, an initial trial compares drug A to drug B, while a different trial studying a similar patient population compares drug B to drug C. Pairwise comparisons fail to define the relationship between drug A and drug C without a trial comparing them directly.

In order to allow indirect comparisons, network meta-analysis was developed as an extension to meta-analysis. It is also called the mixed-treatment comparisons (MTC) method [7]. Compared to meta-analysis, which only allows simple pairwise comparisons, the MTC method can combine all available evidence from a network of trials. Diversity in treatment effects [9] may exist across comparisons in a network, so trials directly comparing drug A and drug C may systematically differ from trials comparing drug A with drug B and drug B with drug C from which an indirect estimate of drug A versus drug C is obtained. MTC models allow to estimate such inconsistency in the evidence structure, and in the absence of considerable inconsistency, to comprehensively estimate effects by taking into account all evidence.

Decision making problems in the context of drug BR analysis often involve multiple criteria (minimum one of each efficacy and safety) and uncertainties (results from clinical trials are estimates with probability distributions). Therefore, only evidence synthesis is not enough, quantifying the trade-offs between benefits and risks under uncertainty is also needed. This can be addressed through Multi-Criteria Decision Analysis (MCDA). MCDA is a discipline aimed at supporting decision makers who are faced with decisions involving multiple criteria [10]. Different MCDA methods allow combining value judgments along multiple dimensions. Multi-Attribute Utility Theory (MAUT) [11] is a traditional MCDA method that assists in the decision making process through maximization of the expected utility. It allows the decision makers (DMs) to quantify the relative importance of the criteria by identifying weights which reflect the DMs’ preferences regarding the criteria’s

relative importance.

Mussen et al. [12] were the first to propose to use MCDA in the context of drug BR analysis. Although they provided a general framework for constructing a multicriteria decision model for BR analysis, their method requires the DMs to provide precise weights for describing the relative importance of criteria. However, in real-life decision problems, it is almost impossible for DMs to provide such exact weights. Felli et al. [13] published a similar MCDA approach in drug BR analysis. They proposed to use categorical value scales for all BR attributes instead of using continuous measurements. Although it simplifies the application of the model, there is a substantial possibility of losing information by mapping measurements from a continuous scale to ordinal categories. Judging from these two applications, it seemed that applying the traditional MAUT couldn't provide a solution for taking into account the uncertainties. In order to overcome the limitations of these two approaches, Tervonen et al. [3] proposed to use Stochastic Multicriteria Acceptability Analysis (SMAA) in drug BR analysis.

SMAA [14] is a family of MCDA methods for supporting multicriteria decisions in situations where neither criteria values nor weights are precisely known. SMAA methods have proved applicable in risk assessment and also other real-life decision problems including Helsinki general cargo harbor EIA [15], elevator planning [16], forest ecosystem management [17], locating an airport hub for centralizing cargo in Morocco [18] and strategic planning of an electricity retailer [19]. Unlike MAUT and other traditional MCDA methods, in SMAA the DMs need not express their preferences explicitly or implicitly and consequently they don't need to know every weight exactly.

Hansen et al. [20] did a meta-analysis for a group of second-generation antidepressants and concluded that they do not differ substantially for treatment of major depressive disorder. Tervonen et al. [3] used a SMAA model to analyze the same antidepressants based on the meta-analysis of Hansen et al. and showed trade-offs between efficacy and safety. However, there still exists a limitation in their work. Although they succeed in quantifying the trade-offs between the benefits and risks, only the evidence from direct comparison

was included in the analysis. The indirect evidence which also contributes to the trade-offs was missing.

In order to overcome this limitation, this thesis evaluates a new approach for drug BR analysis which includes construction of a decision model using SMAA method with criteria measurements from a network meta-analysis. Such model is able to take into account all the available evidence from both direct and indirect comparisons.

In this thesis, the two evaluated approaches will be called the following:

1. “MA/SMAA” represents the old method (meta-analysis based SMAA) proposed by Tervonen et al.
2. “MTC/SMAA” represents the new method (network meta-analysis based SMAA) proposed in this thesis.

The related previous research is described in Table 1.

Table 1: Description of related previous studies and this thesis

Study	Method	Comparison	Criteria	Conclusion
Hansen et al.	meta-analysis	direct	single	No significant difference among drugs.
Tervonen et al.	MA/SMAA	direct	multiple	There exists trade-offs among drugs.
This thesis	MTC/SMAA	direct indirect	multiple	Do the results differ with the one above? Is MTC necessary for drug BR analysis?

## 1.2 Problem statement

This thesis aims to construct the MTC/SMAA model and find out the advantages and disadvantages of this model compared to the MA/SMAA model in the context of drug BR analysis.

## 1.3 Disposition

In the beginning of this thesis, chapter 1, the background and the research problem are described. In chapter 2, the extended background of the research

is introduced in order to let readers understand the theory of the methods and models used in this thesis. Chapter 3 describes the research methodology applied during the research process. The case study is introduced and extended to evaluate MTC/SMAA and compare it with MA/SMAA in chapter 4, and also the analysis and discussion are conducted in this chapter. Finally in chapter 5 the conclusion and recommendation for future work are given. The results from the models are listed in Appendices A-C.

## 2 Extended background

### 2.1 Network meta-analysis

In medical research, meta-analysis is applied frequently to aetiology, to diagnosis and to prognosis, as well as to answer questions of the effects of interventions [21]. This method was initially developed more than a century ago by Karl Pearson [22]. Meta-analysis is a statistical method for combining evidence from multiple studies (often randomized controlled trials) in order to obtain a quantitative synthesis. With the aim of detecting treatment effects reliably, meta-analysis is applied broadly to summarize existing evidence from all available trials as the size of individual trials is often too small. As it tries to estimate a combined effect from a group of individual trials, these individual trials should be similar enough so that the combined estimate is meaningful. However, in practice these trials often vary because of differences in study design. If such variation is excessive, it is called heterogeneity. With meta-analysis one can identify and measure the extent of heterogeneity among the results of trials. As the meta-analysis is based on pairwise comparisons, the results are expressed as ratios. For dichotomous outcomes, effect size is often measured as in risk ratio (RR) or odds ratio (OR). Risk is the probability of having an event, and odds describes how likely an event is to happen, which is the ratio of events to non-events. RR is the risk of the event in one group divided by the risk of the event in the other group. OR is the odds of the even in one group divided by the odds of the event in the other group. OR and RR are similar when the event is rare, but differ substantially when the event is common.

With the arrival of the information era, increasing amounts of information from multiple sources need to be taken into account in analyses that support medical decision making processes. Meta-analysis, which is based on pairwise treatment comparison, starts showing its limitation when dealing with a large network of comparisons. As an example, consider the evidence structures of Figure 1, where the letters represent different drugs and the lines represent direct comparisons between two drugs. In this figure, (1) shows a



pairwise treatment comparison, (2) shows an indirect comparison, (3) shows mixed treatment comparison with both direct and indirect comparison, and (4) shows a realistically complex network of comparisons. The relationship between trials can be arbitrarily complex, but meta-analysis can only access the results from direct comparison between two drugs and fails in drawing conclusions from the indirect comparisons. This causes “selection bias” in the sense that evidence having an indirect relationship with the target will be excluded automatically. Meta-analysis also fails to control the bias which means that it can not generate qualified results if the studies are badly designed. To address this situation, there exists an extension to the meta-analysis which allows multiple (more than two) treatments to be simultaneously compared. This method is called network meta-analysis or mixed treatment comparison (MTC) [7]. In MTC models, all available evidence can be taken into account from both direct and indirect comparisons.

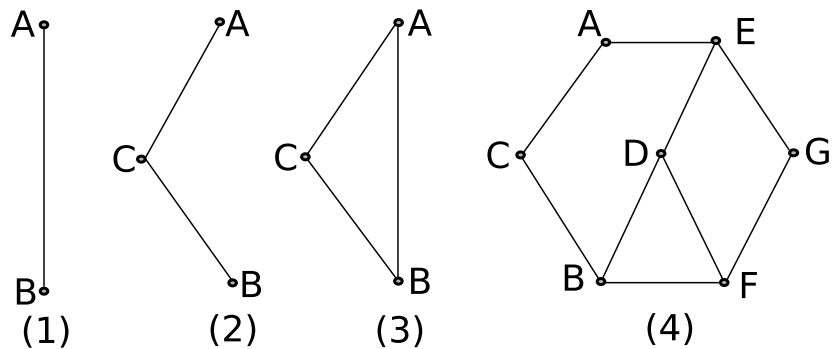


Figure 1: Network evidence structures.

The main contributions of the MTC method are that it includes all the available evidence, it allows checking the consistency of the evidence formally [23], and that it does not depend on the chosen comparator treatment. In addition, the MTC method improves the transparency of evidence synthesis process as treatments don't need to be grouped into multiple analyses or excluded. Evidence synthesis through the MTC method provides a basis for single-criterion decision making.

### 2.1.1 Consistency model

Suppose we have three treatments: A, B and C, as shown in Figure 1 (3). The three pairwise comparisons are AB, AC and BC. We derive three unrelated estimates of the treatment effects  $d_{AB}$ ,  $d_{AC}$ ,  $d_{BC}$  from the given studies by carrying out three separate meta-analyses, which measure the treatment effects on a scale symmetric at zero [9], such as Log-OR. The consistency model assumes a transitivity among the three parameters, so their relationship can be described as follows:

$$d_{BC} = d_{AC} - d_{AB} \quad (1)$$

The assumption of consistency states that the parameter  $d_{BC}$  can be estimated using both direct BC comparison evidence and indirect comparison evidence on AC and AB. In a consistency model, direct and indirect evidence is clearly related. Under consistency, a Bayesian random-effects model can be applied to estimate the treatment effect [8].

### 2.1.2 Inconsistency model

In order to assess inconsistency, an additional term is used to represent the inconsistency [9], i.e.:

$$d_{BC} = d_{AC} - d_{AB} + \phi \quad (2)$$

where  $\phi$  represents the inconsistency between the BC treatment effect from direct BC trial and the BC treatment effect based on indirect evidence from AB and AC trials. For example, if  $d_{BC}$  measured from direct BC trial is 0.5 and  $d_{AC}$  and  $d_{AB}$  from trials AC and AB are 0.7 and 0.3, the indirect comparison of the treatment effect between B and C is 0.4. This then is inconsistent with 0.5 from the direct comparison, and therefore  $\phi = -0.1$ . In general, the inconsistency can be expressed as a vector:

$$\phi = (\phi_1, \phi_2, \dots, \phi_L) \quad (3)$$

When  $\phi=0$ , the model is a consistency model. When  $\phi\neq 0$ , it is an inconsistency model. The dimension  $L$  of  $\phi$  describes the inconsistency degrees of freedom (ICDF). Normally it can be identified as the number of independent loops in the graph representation of the evidence network. If all the trials are two-arm trials,  $L$  can be calculated by  $L=J-(T-1)$ , where  $J$  is the number of different pairwise comparisons and  $T$  is the number of treatments. Take Figure 1 (4) as an example, where  $T=7$ ,  $J=9$ , so  $L=3$ . If some trials have more than two arms, the calculation of  $L$  is more complicated [24].

The inconsistency of a network could reflect genuine diversity, bias or a combination of both [9]. The lack of a demonstrable inconsistency does not prove that the results are free of bias and diversity. It is particularly difficult in clinical or epidemiological areas to evaluate the inconsistency, because important characteristics may not be reported, many comparisons can have few results or the results are reported in diverse formats.

The MTC method allows estimation of both heterogeneity in the effect of any given treatment and inconsistency in the evidence from different pairs of treatments [7]. Assume that drug A is found to be superior to drug B in the first trial, and drug B to be equivalent to drug C in a second trial. Network meta-analysis then allows one to say that drug A is also superior to drug C for this particular patient population.

### 2.1.3 Implementation of MTC

Network meta-analysis can model various types of outcome data through different probability distributions. For example, a binomial distribution can be used to model numbers of events, or a normal distribution to model sample means [9]. The MTC model is implemented as a Bayesian estimation procedure through sufficiently flexible software, such as WinBUGS or JAGS, which uses Markov Chain Monte Carlo simulation [8, 9]. For more details on algorithmic parametrization of MTC models, see [24].

## 2.2 Stochastic Multicriteria Acceptability Analysis

SMAA methods have been developed for supporting discrete group decision making problems in situations where neither DMs' preferences nor criteria measurements are precisely known. Such decision making problems occur often in practice. According to Lahdelma et al. [25] the reasons why the preference information is often difficult to obtain include DMs not having sufficient time to study the problem or them having difficulties in comparing criteria so that they cannot provide precise preference information, and the DMs being afraid of revealing their preferences in public. The reason could also be that the DMs do not want to fix their their preferences because they may change during the process, or that the analysts are unable to elicit the preferences. However, even if the preference information can be obtained from the DMs, the preferences of several DMs that disagree are hard to combine and any subjective weight information should be considered uncertain since different weight elicitation methods provide different weights for the same problem [25]. Instead of trying to identify the best alternative based on precise preference information [25], SMAA allows to identify the preferences that make each alternative the preferred one. The fundamental idea of SMAA is to quantify the decision uncertainty and to aid decision making through descriptive measures calculated as multidimensional integrals over stochastic parameter spaces [26].

There exists a number of different variants of SMAA. SMAA-2 [27] extends the original SMAA method to consider any rank from best to worst for each alternative. SMAA-3 [28] is based on pseudocriteria as in the ELECTRE III decision-aid. SMAA-III [29] applies the full ELECTRE III outranking process with uncertain criteria measurements, weights, and thresholds. SMAA-D [30] uses, instead of a value function, the efficiency score of Data Envelopment Analysis. SMAA-P [31] is based on piecewise linear prospect theory where alternatives are evaluated with respect to gains and losses from reference points. SMAA-O [25] provides a solution to treat mixed cardinal and ordinal criteria data. SMAA-A [32] compares the alternatives by applying reference points and achievement functions. SMAA-TRI [33] is an ordinal classifica-

tion method based on ELECTRE-TRI with uncertain criteria, thresholds, and weights.

### 2.2.1 SMAA-2

The SMAA-2 method [27] is chosen in this thesis as it allows directly ranking all the alternatives. It explores the multi-dimensional weight space based on an assumed value function and stochastic criteria values, in which preferences and uncertainties are quantified and arbitrarily distributed criteria measurements can be used.

Consider that a group of DMs have a set of  $m$  alternatives  $\{x_1, x_2, \dots, x_m\}$ , which are evaluated in terms of  $n$  criteria. A real-valued utility function  $u(x_i, w)$  is assumed to represent the DMs' preference structure, where  $w$  is an individual weight vector for a DM to quantify their subjective preferences. The weights are assumed to be non-negative and normalized, and the feasible weight space  $W$  is defined as

$$W = \left\{ w \in R^n : w \geq 0 \text{ and } \sum_{j=1}^n w_j = 1 \right\} \quad (4)$$

The uncertain or imprecise criteria values are represented by stochastic variables  $\xi_{ij}$  corresponding to assumed or estimated joint probability distribution and density function  $f(\xi)$ , where  $i \in \{1, \dots, m\}$  index alternatives and  $j \in \{1, \dots, n\}$  index criteria. Similarly, the DMs' unknown or partially known preferences are represented by a weight distribution with density function  $f(w)$  in the feasible weight space  $W$ . Total lack of DMs' preference information is represented by a uniform weight distribution in  $W$ :

$$f(w) = 1/\text{vol}(W) \quad (5)$$

where the  $(n-1)$ -dimensional volume of the feasible weight simplex is

$$\text{vol}(W) = n^{1/2} / (n-1)! \quad (6)$$

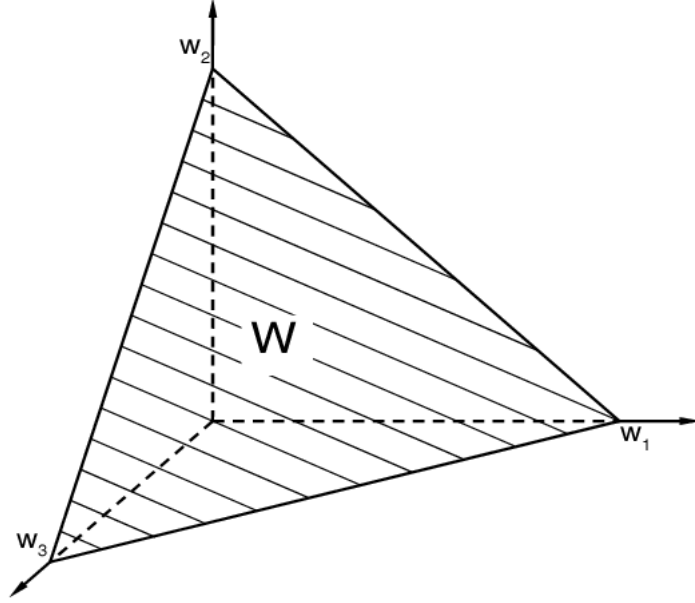


Figure 2: The feasible weight space in the three-criteria case.

In the three-criteria case  $W$  is a two-dimensional simplex as illustrated in Figure 2.

If criteria values and weights are precisely known, the problem can be easily solved by evaluating the utility function for each alternative and choosing the one with largest utility. However, normally neither criteria values nor weights are precisely known in real-life decision-making problems. The rank of each alternative is defined as an integer from the best rank (1) to the worst rank ( $m$ ) by means of a rank function,

$$rank(\xi_i, \omega) = 1 + \sum \rho(u(\xi_k, \omega) > u(\xi_i, \omega)) \quad (7)$$

where  $\rho(true) = 1$  and  $\rho(false) = 0$ . The set of favorable rank weights  $W_i^r(\xi)$  for each alternative are defined as

$$W_i^r(\xi) = \{w \in W : rank(\xi_i, \omega) = r\} \quad (8)$$

For aiding the decision making, SMAA-2 defines three descriptive measures,

the rank acceptability index, the central weight vector, and the confidence factor.

**Rank acceptability index  $b_i^r$**  The rank acceptability index is a measure of the variety of different valuations making an alternative the most preferred one, and also takes into account the acceptability for a certain rank. The rank acceptability index measures the variety of different values granting alternative  $x_i$  rank  $r$ . It is computed as an integral over the criteria distributions and the favorable rank weights as

$$b_i^r = \int_X f(\xi) \int_{W_i^r(\xi)} f(w) dw d\xi \quad (9)$$

Normally the most acceptable alternative should be the one with the highest acceptability for the best rank. The rank acceptabilities can be used directly in the multicriteria evaluation of the alternatives [27].

**Central weight vector  $w_i^c$**  The central weight vector is defined as the expected center of gravity of the favorable weight space. It is computed as an integral over the criteria and weight distributions by

$$w_i^c = \int_x f(\xi) \int_{w_i^r(\xi)} f(w) w dw d\xi / b_i^1 \quad (10)$$

The central weight vectors represent the typical valuations resulting in the decision [14], and they can be presented to the DMs to help them understand how different weights correspond to different choices [27]. Under such circumstance, the DMs can know which actions should result from what preferences without providing any preference information.

**Confidence factor  $p_i^c$**  The confidence factor is defined as the probability for an alternative to be the preferred one with the chosen central weight vector. It is computed as an integral over the criteria distributions as

$$p_i^c = \int_{\xi \in u(\xi_i, w_i^c) \geq u(\xi_k, w_k^c)} f(\xi) d\xi \quad (11)$$

Confidence factors measure whether the input data is accurate enough for making an informed decision [14]. They quantify whether a decision is certain when the DMs' preferences correspond to certain central weights.

As mentioned above, SMAA is developed for the situations where neither the criteria nor the preferences are precisely known. However, in some decision problems, it is possible to obtain some specific weight information. Under such circumstance, the density functions is defined with a uniform distribution in the restricted weight space  $W'$  as  $f'(w) = 1/vol(W')$  when  $w \in W'$  and  $f'(w) = 0$  when  $w \in W \setminus W'$ . In particular, SMAA-2 introduces the following types of restrictions on the weight space [27]:

1. Partial or complete ranking of criteria ( $w_j \geq w_k$ ).
2. Intervals for weights ( $w_j \in [w_j^{min}, w_j^{max}]$ ).
3. Intervals for weight ratios (trade-offs) ( $w_j/w_k \in [w_{jk}^{min}, w_{jk}^{max}]$ ).
4. Linear inequality constraints for weights ( $Aw \leq c$ ).
5. Nonlinear inequality constraints for weights ( $f(w) \leq 0$ ).

The feasible weight space  $W$  with interval constraints for weight  $w_1$  in the three-criteria case is shown in Figure 3. A complete ranking can be obtained by asking the DMs to identify the most important, second important, etc. criterion. The complete ranking of the criteria can be expressed as a sequence of inequality constraints of the weights

$$w_{j1} > w_{j2} > \dots > w_{jn} \quad (12)$$

The feasible weight space  $W$  in the three-criteria case is showed in Figure 4 with the ranking  $w_1 > w_2 > w_3$ . The preference information from multiple DMs can be combined using union, intersection, or averaged density [27].



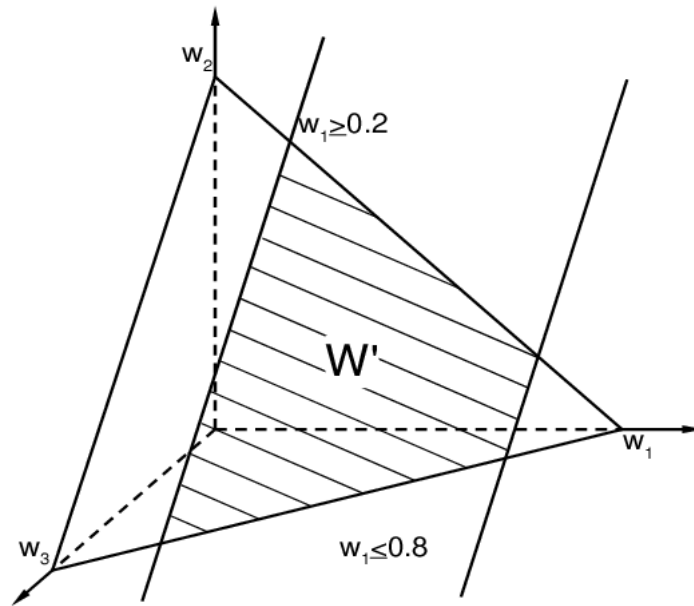


Figure 3: The feasible weight with constraints on  $w_1$  in the three-criteria case.

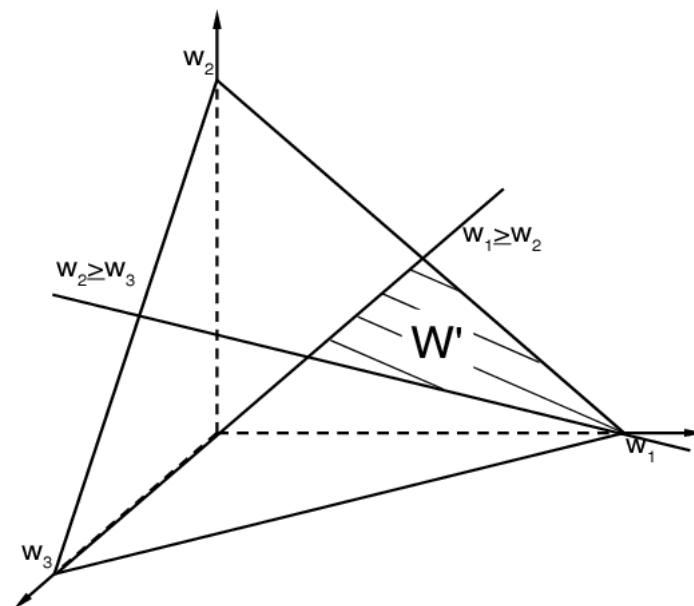


Figure 4: The feasible weight space with complete ranking of the weights in the three-criteria case.

From Figures 3 and 4 we can see that the preference information limits the size of weight space. The central weights are generally considered only if there is no preference information included in the model, as their descriptive power is limited in case of being computed within a restricted weight space.

### 2.2.2 Implementation of SMAA

SMAA computations are based on multidimensional integrals, which in practice are impossible to calculate manually because the distributions can be arbitrarily complex. Straightforward integration techniques are also not feasible because they are based on exploring the distributions with respect to each dimension and the required effort depends exponentially on the number of dimensions. However, Monte Carlo simulation provides a proper solution because a very high precision answer is not necessary. Monte Carlo simulation is widely used in mathematics for the evaluation of definite integrals, and it is particularly useful for multidimensional integrals with complicated boundary conditions.

The computation of SMAA algorithm is done in four parts, which are the generation of the criteria measurement matrix, the generation of weights, the computation of  $b_i^c$  and  $w_i^c$ , and the computation of  $p_i^c$ . For more details about the algorithm, see [34].

## 3 Research methodology

### 3.1 Research type

In order to address the research problem stated in section 1.2, different approaches were considered to perform the research at the beginning.

First alternative is to perform qualitative comparison between the MA/SMAA and MTC/SMAA models. The qualitative comparison can provide an in-depth understanding of the two models and draw the conclusion by comparing the principle of construction and implementation processes of the two models. The advantages and disadvantages of the MTC/SMAA model are mainly according to the comparison between MTC model and meta-analysis as discussed in section 2.1. The strength of this approach is that the comparison is based on the model structure design, free of data and gram processing, and the conclusions are not limited in any specific case or area. However, there is no verification to support such conclusions. For instance, one advantage of the MTC model is that it allows taking into account all available evidence, but how this allowance will influence the results from SMAA model is not illustrated, e.g. whether the results will assist the decision making or not influence the results at all, etc. Therefore the general conclusions produced through such qualitative comparison are only hypotheses, and evidence to verify the hypotheses is missing [35].

Second alternative is to study multiple cases and use quantitative analysis for concluding which model outperforms the other. The case study method is chosen as it is based on in-depth investigations and provides a systematic way of looking at the case, collecting and analyzing the data, and reporting the results [36]. With case studies, the hypotheses can be generated and verified by themselves [37]. The MTC/SMAA model is constructed and evaluated through case studies so that the construction and implementation processes are shown clearly, and the results will be supported by large amount of data with which a general conclusion can be drawn through quantitative analysis. The strength of this approach is that the applicability of the models can be tested through different cases, and quantitative analysis concludes the

advantages and disadvantages of the MTC/SMAA model from the results viewpoint with large amount of data supporting. However, unfortunately due to limited suitable case data in drug BR analysis area, studying multiple cases is not realistic. In addition, quantitative analysis mainly draw the conclusion based on observation of these empirical evidences from multiple cases on the effectiveness of the two models, but is unable to provide a sufficient comparison on the models structure design and their functionality which are important for how the DMs apply the models.

Considering the pros and cons of the two alternatives discussed above and properties of research problem, another method is applied, which is studying a single case and concluding the advantages and disadvantages of the MTC/SMAA model by both quantitative and qualitative comparisons with the MA/SMAA model. The property of research problem in this thesis, which is the need of a specific scenario to develop and analyze the models so that the results are comprehensible and measurable, decides case study method. Due to limited suitable data resource and the fact that the MA/SMAA model was only applied in one case, the same case study, efficacy and safety analysis of second-generation antidepressants, will be applied and extended in this thesis. Using the same case allows the comparison of the MA/SMAA and MTC/SMAA model under the same criteria and scenarios. In order to draw a conclusion based on the comparison between the two models, both quantitative and qualitative comparisons are applied. Quantitative comparison is performed on the results (parameters) from the models as it evaluates the effectiveness of the models. Qualitative comparison is performed from the model structure design and implementation requirements point of view as it evaluates the functionality and complexity of the models. The strength of this method is that the MTC/SMAA model is applied in the same case with the MA/SMAA model which allows a feasible and fair comparison between them, and the results from the MA/SMAA model already existed. In addition, combination of quantitative and qualitative comparisons allows a holistic perspective analysis and also a verification supporting the hypotheses made in the qualitative comparison. However, the limitation of this method is that the applicability of the MTC/SMAA model in other cases or areas is

unknown.

## 3.2 Research steps

The research is conducted with several steps including literature review, data collection, model construction, analysis, and concluding. During these steps, different methods could be applied, and each chosen method is motivated by the specific properties and requirements of the research problem.

**Literature review** This thesis requires knowledge from cross-disciplines of information systems, decision sciences, statistics and pharmacology, so large amount of literature needs to be reviewed. This literature review mainly covers areas of drug information system, MCDA, SMAA, meta-analysis, network meta-analysis, decision making, BR analysis, and Bayesian analysis. Most of the literature is obtained from the digital library at KTH and RuG.

**Data collection** Data collection is the process of gathering related information to identify this specific topic [38]. Different methods could be applied to collect data, such as personal interviewing, observing, extracting available information, searching Internet, group discussion, questionnaires, etc. [38]. In this thesis, the required data is the information of treatments with depressive patients, such as study groups, patient characteristics, number of treatment responders, number of dropout patients, number of patients who get adverse drug events. The number of treatment responders and number of patients who get adverse drug events are most important as they generate the criteria measurements to efficacy and safety of the drugs (see the detailed definition in section 4.1.1 and 4.1.2). The data mostly comes from published clinical trials results in literature, therefore the approach extracting available information is the only feasible way to collect data in this thesis. The data was extracted from 44 original studies, listed in Appendix D, comparing different antidepressants which were cited by Hansen et al. [20].

**Model construction** The MTC/SMAA models were constructed in the chosen case study: first a network meta-analysis was performed to synthesize the evidence, and then a SMAA decision model was constructed using its results. Limited knowledge and technology results in two separate steps in the MTC/SMAA model generation, the MTC model and the SMAA model. The MTC model can be constructed manually or by using specific software. Due to the complexity of the model construction, in this thesis “drugis.org MTC library” [24] is chosen to generate the models, JAGS [39] is used to calculate the models, and the post-processing to transform the results into a comprehensible format is conducted through R [40]. The SMAA model can be constructed by hand or generated automatically, and in this thesis it was built up by inputting the parameters manually into JSMAA [41] as those parameters are generated from another separated model, the MTC model.

**Analysis** In this thesis, the analysis focuses on comparing the MTC/SMAA and MA/SMAA models so that the advantages and disadvantages of the former are evaluated. As discussed in section 3.1, combination of the quantitative and the qualitative comparisons allows a systematic method to draw the conclusion and also provides an analysis of the result from multiple perspectives. With the analysis of results the following questions are answered:

1. Do the SMAA decision analysis results differ significantly with the criteria measurements are constructed from network meta-analysis from those that are constructed from meta-analysis?
2. How do the two models differ qualitatively from the decision support viewpoint?
3. How easily can the two models be implemented as alternative evidence synthesis options in a decision support system?

To answer the first question, the quantitative comparison is conducted from the results perspective, which refers to the performance of each antidepressant under different scenarios. The corresponding outputs from the two methods are compared by estimated effects, rank acceptabilities, central weights

and confidence factors, which are the parameters of shown results from the models. These criteria are chosen as not only they are the indices of the decision model results but also they reflect the outcome of incorporating all available evidence (by estimated effects), which model produces more discriminative results (by rank acceptabilities), and whether the trade-offs between benefit and risk are identified (by central weights and confidence factors). The quantitative comparison verifies the hypotheses made in the qualitative comparison and deals with how the DMs can make decisions by applying the models. However, the scales for different measurements in the two models should be consistent to perform the quantitative comparison so that those numbers have the same meaning, as otherwise different results will not tell much about the differences between the two models.

To answer the rest two questions, the qualitative comparison is processed from the requirements perspective which refers to what is required to implement the method, what is the difference in the complexity of the methods and what additional information is required in application of the MTC/SMAA model for pharmacological decision making. The complexity is measured by analyzing the steps required to implement the method, model generation and format of the results. These criteria tell the difference between the two models with respect to how easily the DMs can understand and apply the models. Through the qualitative comparison the reasons for the differences in the quantitative comparison are explained.

**Concluding** The MTC/SMAA model is constructed and evaluated in the context of drug BR analysis. The evaluation is processed according to the results from quantitative and qualitative analyses. The conclusion is drawn by the answers to the three questions in analysis step.

The first question (with quantitative comparison) is answered by the discrimination of the results, where the term “discrimination” refers to the difference between each drug’s rank acceptability in the best and worst rank and the scale of one drug outperforms the others in a certain scenario. Under the situation that the parameters are generated from available evidence correctly,

the more discriminative the result is, the easier for DMs to make the decision based on the result. However, to treat “discriminative” as an advantage or disadvantage of the MTC/SMAA model is not suitable for all the situation. For example, in the case chosen by this thesis, more discriminative results mean better results from decision making model, but sometimes in other cases it is not guaranteed. Therefore, in the conclusion, discrimination is considered as a criterion to evaluate the applicability of the MTC/SMAA model in this specific case, but not as the advantage or disadvantage of the model in a general.

The second question (with qualitative comparison) is answered by indicating the transparency of the decision making process, where the term “transparency” refers to dealing with those under-table information, such as the inconsistency and the automatic excluded studies or evidence. Also this question is answered by the functionality of the models which refers to their ability of synthesizing evidence and supporting the decision making.

The third question (with qualitative comparison) is answered by the complexity of using the models for DMs, where the term “complexity” refers to how easily can DMs understand and implement the models.

The recommendation for the future work is made according to the observation during the research and the limitation of applied method.



## 4 Case study

In order to go beyond pure theoretical discussion, a real case, efficacy and safety analysis on second-generation antidepressants, is introduced and analyzed with the MTC/SMAA method. The whole process of case study is presented in Figure 5. The results from the MA/SMAA method are listed in Appendix A (adapted from [3]), and the results from the MTC/SMAA method are listed in Appendix B.

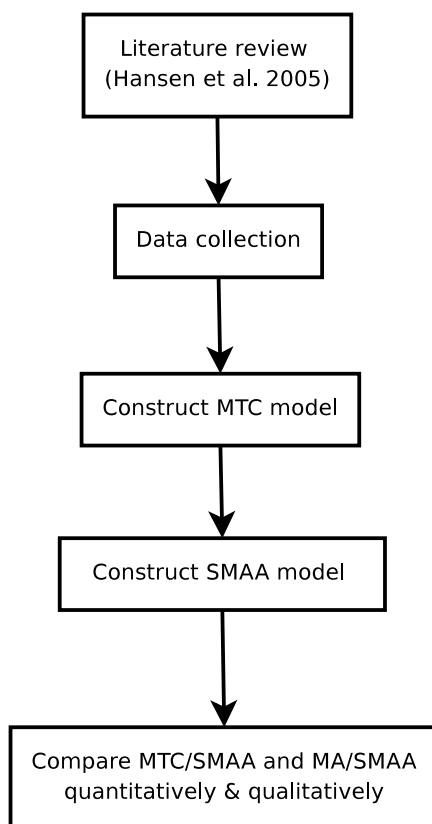


Figure 5: Process diagram of case study.

Forty-four head-to-head randomized, controlled trials compared ten commonly prescribed second-generation antidepressants. The network of all available studies is depicted in Figure 6, where each line indicates the direct comparison between two drugs. The number close to each line indicates the number of studies which include this comparison, and the dotted circle

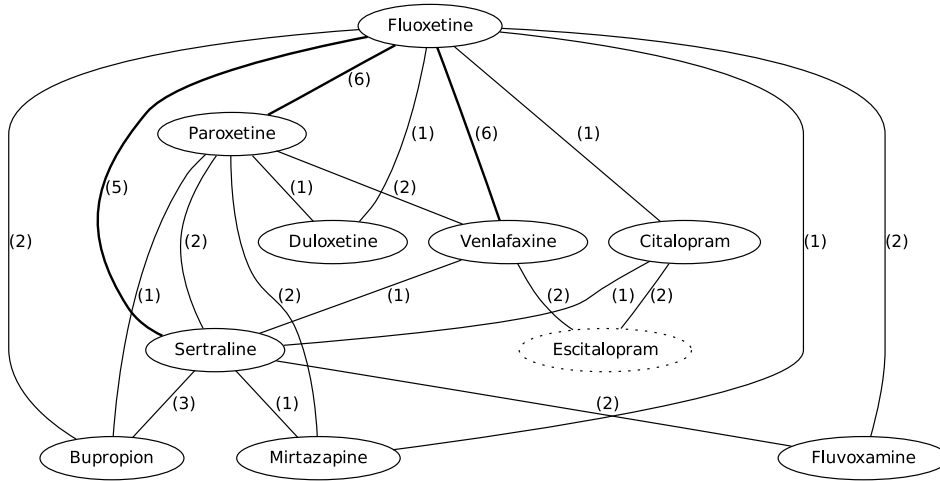


Figure 6: Evidence network of existing studies comparing efficacy of second-generation antidepressants.

around Escitalopram means that there is no direct comparison between it and Fluoxetine which is chosen as the common comparator. Since meta-analysis is based on pairwise comparisons, a common comparator is required to perform the analysis. Figure 6 shows that only 20 comparisons are available out of 45  $(n(n-1)/2, n=10)$  possible comparisons in the included studies.

## 4.1 Previous related studies

### 4.1.1 Meta-analysis

Meta-analysis has been applied in a published treatments comparison by Hansen et al. [20]. They reviewed and analyzed the 44 studies comparing second-generation antidepressants and performed 3 meta-analyses (the bold lines in Figure 6) on Paroxetine, Sertraline and Venlafaxine with Fluoxetine as the baseline. In these meta-analyses, 5 studies among Sertraline, Paroxetine and Venlafaxine were excluded (see Figure 7). They used treatment response as the benefit criterion to evaluate efficacy, defined as 50% or greater improvement on the Hamilton Rating Scale for Depression (HAM-D) or the Montgomery-Asberg Depression Rating Scale (MADRS) from baseline to the endpoint.

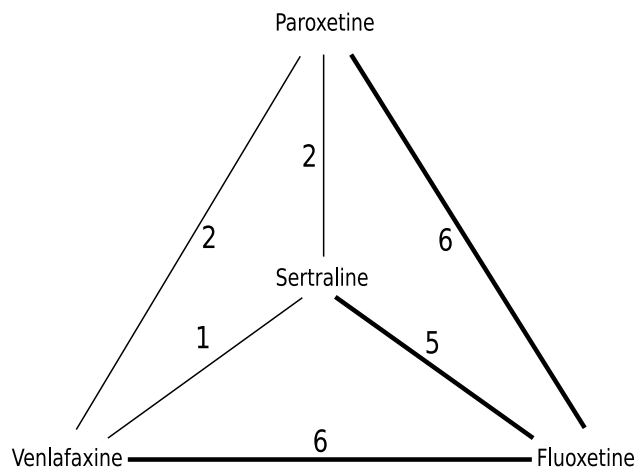


Figure 7: Evidence network of chosen antidepressants.

Hansen et al. concluded that the four antidepressants do not differ substantially for treatment of major depressive disorder. For example, the meta-analysis of Fluoxetine compared with Paroxetine is described by a forest plot, shown in Figure 8 (adapted from [20]). Five studies comparing Fluoxetine with Paroxetine were pooled. The numbers on each side of the 95% confidence intervals (CI) are the number of responders over the total number of participants who were randomly allocated to receive that drug, and the relative benefit is a risk ratio. The forest plot tells us that the combined relative benefit overlaps with 1, which means that there is no significant difference between Paroxetine and Fluoxetine (relative benefit, 1.09 [CI, 0.97 to 1.21]) [20].

#### 4.1.2 MA/SMAA

Based on the safety and efficacy data from the meta-analysis done by Hansen et al., Tervonen et al. [3] constructed a SMAA-2 model for the therapeutic group of antidepressants. In their study, four antidepressants for which sufficient quantitative data was available were selected for the analysis: Fluoxetine, Paroxetine, Sertraline, and Venlafaxine. They used the same efficacy criterion as in Hansen et al. [20]. The risk criteria to evaluate safety were the five most common adverse drug reactions (ADRs): diarrhea, dizziness,

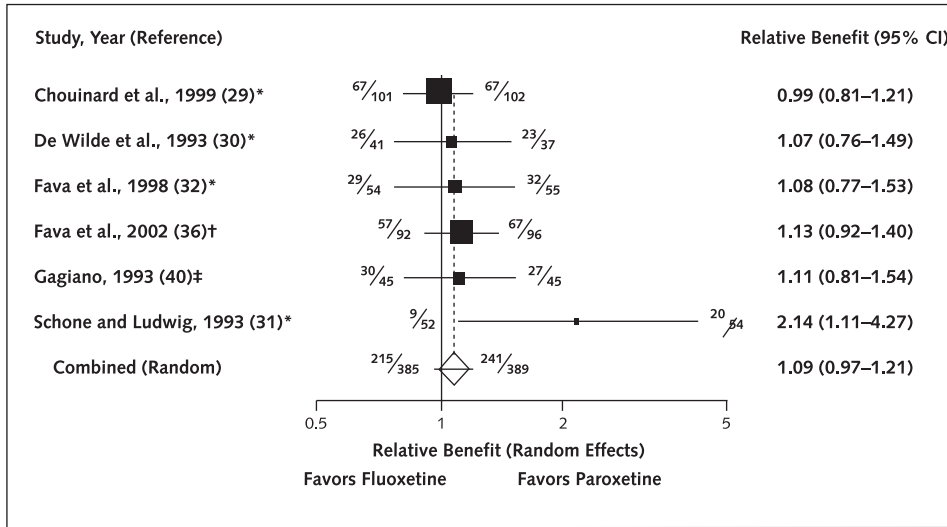


Figure 8: Meta-analysis of Fluoxetine compared with Paroxetine.

headache, insomnia, and nausea. The criteria are summarized in Appendix A Table 4. The efficacy and safety data from the meta-analysis is shown in Appendix A Table 5.

Tervonen et al. used the meta-analyses of Hansen et al. as estimates of the relative efficacy of Paroxetine over Fluoxetine, Sertraline over Fluoxetine, and Venlafaxine over Fluoxetine. The pooled ADRs were also taken from the published results of Hansen et al. The pooled incidences of the ADRs as well as the log of the pooled efficacy ratios were assumed to be independently and normally distributed random variables [42]. The means of these distributions are equal to the (log of the) pooled effect size estimates, and the standard deviations  $\sigma$  are derived from the corresponding 95% CI as reported in the meta-analysis [3]. For example, the pooled dizziness for Paroxetine was found to be equal to 10.6 with a 95% CI of [7.5-13.7] [20]. The estimated standard deviation of the pooled incidence of dizziness for Paroxetine can be calculated as  $\sigma_{dizziness}^{Paroxetine} = \frac{13.7-7.5}{2*1.96} = 1.58$  as the upper (lower) bound of this CI was computed by adding (subtracting)  $1.96\sigma$  to (from) the effect-size estimates of 10.6.

The analyses were performed under three scenarios: one with missing preference information, and two with a criteria ranking elicited from an expert

in the field of antidepressants: separately for mild and severe depression. The first scenario was with no preference information. Its rank acceptability indices are listed in Appendix A Table 6 and visualized as a column chart in Appendix A Figure 14. The central weights are listed in Appendix A Table 7. These results indicate that all drugs have reasonable BR profiles and there are clear trade-offs among the four drugs. For instance, if the DMs have a prior preference for Paroxetine, then according to the BR profiles expressed through the central weights, apparently the ADR nausea displays the highest importance. To make a rational decision, the DMs should favor lowering first nausea from the worst scale value (34%) to the best scale value (11.1%).

In the remaining two scenarios, the strict preference relation is denoted by  $\succ$ , where “ $A \succ B$ ” means “ $A$  is preferred to  $B$ ”. The preference ranking for mild depression is Diarrhea  $\succ$  Nausea  $\succ$  Dizziness  $\succ$  Insomnia  $\succ$  Headache  $\succ$  Efficacy, and for severe depression is Efficacy  $\succ$  Diarrhea  $\succ$  Nausea  $\succ$  Dizziness  $\succ$  Insomnia  $\succ$  Headache <sup>1</sup>. The rank acceptability indices for the scenario of mild (severe) depression are listed in Appendix A Table 8 (Appendix A Table 9) and visualized as column charts in Appendix A Figure 15 (Appendix A Figure 16). From the figures we can see that both the mild and severe depression scenarios result in a relatively high first rank acceptability for Paroxetine. Since it also has a good rank in the first scenario, it can be considered to have the “best” overall BR profile if no additional information is available. The rank acceptabilities for other drugs are more sensitive to the preference information.

## 4.2 MTC/SMAA

To evaluate the MTC/SMAA method, a multicriteria model is constructed for the same group of second-generation antidepressants to demonstrate the applicability of the MTC/SMAA model in drug BR analysis, as described in section 3.2. The same criteria are chosen as in the study of Tervonen et al. [3] so that the MTC/SMAA method is processed under the same scenario with

---

<sup>1</sup>The ranking was provided by an expert in the field of antidepressants for mild depression and severe depression.

the MA/SMAA method. The criteria characteristics are listed in Appendix B Table 10.

#### 4.2.1 MTC model

In the network meta-analysis part, as shown in Figure 9, at first an inconsistency MTC model is constructed, and its results are checked for a significant amount of inconsistency. If there is no relevant inconsistency, the consistency MTC model can be constructed which will be then used for constructing the SMAA model. If there is inconsistency, the reason for the inconsistency must be determined. Then, the explanation of inconsistency will be judged by medical experts. If the explanation is sufficient to identify the source of inconsistency, the causing study is removed, and a new inconsistency MTC model is constructed and inconsistency evaluation is repeated.

To make the whole network meta-analysis clear, consider an imaginary analysis of three second-generation antidepressants, Fluoxetine (F), Paroxetine (P) and Venlafaxine (V). Assume that there are 8 studies comparing F and P, 10 studies comparing F and V, and 5 studies comparing P and V (see Figure 10). The inconsistency model shows that the result from direct comparison V-P is inconsistent with the result from indirect comparison V-P through comparisons F-P and F-V. After the judgment from medical experts, the reason for the inconsistency is that in the studies comparing F-V and F-P, all the patients are female, but in two of the studies comparing V-P, the patients they research on are male, so these two studies are removed. This step is done repeatedly until there is no relevant inconsistency, and then the consistency model is constructed.

The analysis with Fluoxetine, Paroxetine, Sertraline, and Venlafaxine is conducted using Fluoxetine as the baseline. The criteria measurements from the network meta-analysis are listed in Appendix B Table 11. Markov Chain Monte Carlo simulation with 50,000 iterations is used to estimate each MTC model. The results of a consistency MTC model <sup>2</sup> for the efficacy (HAM-D)

---

<sup>2</sup>In this thesis, the inconsistency model is not shown as the analysis of inconsistency is

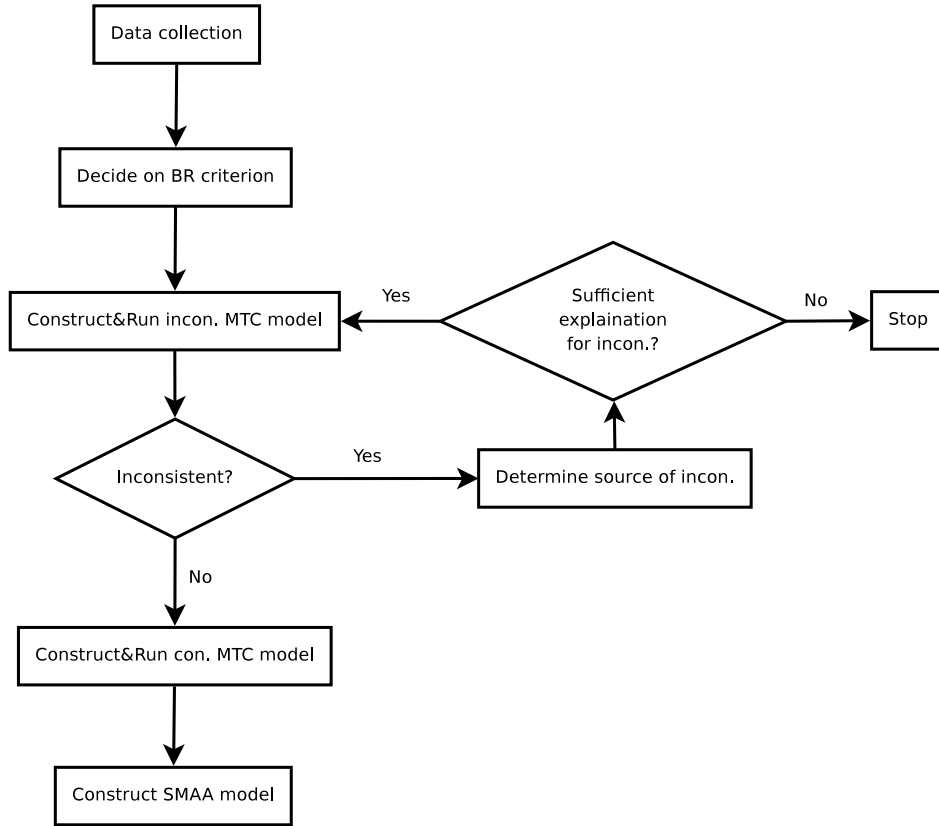


Figure 9: Diagram of MTC model application.

criterion are shown in Table 2. Each cell gives the OR (95% CI), where the row-defining treatment  $i$  is baseline and the column-defining treatment  $j$  the comparator:  $\frac{odds(j)}{odds(i)}$ . ORs higher than one indicate a higher response-rate for the column-defining treatment. In order to input these measurements into JSMAA, the results are transformed to Log-OR ( $\pm$ SD), where SD stands for Standard Deviation. For example, the OR (95% CI) for Paroxetine compared to Fluoxetine is 1.20 (0.91,1.58), so the corresponding Log-OR ( $\pm$ SD) is 0.17657 ( $\pm$ 0.13745), which can be used for the SMAA model.

The criteria measurements in the MTC/SMAA model are Log-OR relative to Fluoxetine, which are meaningless to DMs. Therefore the result is transformed into an absolute risk value (except for efficacy) in order to have similar

---

complicated and requires medical domain knowledge. The result of inconsistency analysis for this case is that there is no significant inconsistency in this model.

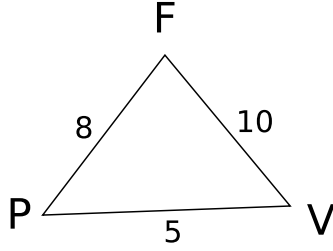


Figure 10: An imaginary example of a study network.

Table 2: Consistency model results

Fluox	1.20 (0.91, 1.58)	1.27 (1.00, 1.62)	1.42 (1.14, 1.76)
0.84 (0.63, 1.10)	Parox	1.06 (0.75, 1.50)	1.18 (0.87, 1.61)
0.78 (0.62, 1.00)	0.94 (0.67, 1.32)	Sertr	1.11 (0.82, 1.51)
0.71 (0.57, 0.88)	0.84 (0.62, 1.15)	0.90 (0.66, 1.22)	Venla

scale units with the MA/SMAA analysis. The MTC/SMAA results under absolute measurements are listed in Appendix C.

#### 4.2.2 SMAA model

For the SMAA analysis, three scenarios are processed: no preference information, mild depression, and severe depression. In the first scenario, no preference information, the rank acceptability indices are listed in Appendix C Table 18 and visualized as a column chart in Appendix C Figure 20. The central weights are listed in Appendix C Table 19. From the central weights, clear trade-offs among the four drugs are identified. For instance, if the DMs display a higher preference of Fluoxetine, apparently from Appendix C Table 19 we can see that nausea has the highest relative importance. Then if this model is accepted by DMs, they should favor lowering first nausea from the worst scale value (40.89) to the best scale value (13.81). By looking at DMs' preferences for scale swings (Appendix C Table 16) with the central weights (Appendix C Table 19), the DMs can decide the preferable drug under specific circumstances. For example, if the DMs consider the scale swing



of efficacy (0.86) more important than the scale swing of headache (27.64), then Fluoxetine should be out of their preference compared to other drugs as it is the only drug whose central weight of efficacy is considerable lower than the central weight of headache.

In the remaining two scenarios, rank acceptability indices for the scenario of mild (severe) depression are listed in Appendix B Table 14 (Appendix B Table 15) and visualized as column charts in Appendix B Figure 18 (Appendix B Figure 19). From the figure we can see that in the mild depression scenario, Fluoxetine has considerably higher first rank acceptability and much lower last rank acceptability than the others. This result can already give DMs the supportive information that in the mild depression scenario Fluoxetine should be preferred over the other three drugs if no additional information is available. In the severe depression scenario, Venlafaxine performs better compared to the rest. In addition, compared to Paroxetine and Sertraline, the rank acceptabilities of Fluoxetine and Venlafaxine are much more sensitive to the preferences, since their rank profiles completely depend on preferences. In the mild depression scenario Fluoxetine achieves a significantly higher first rank acceptability (0.77) than in the severe depression scenario (0.03). The situation is completely opposite with Venlafaxine: its first rank acceptability in severe depression scenario (0.49) is much higher than in the mild depression scenario (0.02).

## **4.3 Analysis**

### **4.3.1 Quantitative comparison**

Before conducting the quantitative comparison between the MA/SMAA model and the MTC/SMAA model, the measurements in the MTC/SMAA model need to be modified in order to be consistent with the MA/SMAA model. In the MTC/SMAA model, the incidences of ADRs are measured in relative value (Log-OR) to Fluoxetine, which are hardly meaningful to DMs as it is difficult to interpret them without knowledge of the risk of ADRs incidences with Fluoxetine. Therefore, after the absolute odds for Fluoxetine are

assumed, the results are transformed into absolute risk value using the sampled odds as  $\text{risk} = \text{odds}/(1+\text{odds})$ . Tervonen et al. [3] also used absolute risk measurements for ADRs in the MA/SMAA model, and the conversion of MTC/SMAA ADRs measurements to this scale is necessary to perform a fair comparison and to reuse the criteria rankings.

By contrasting the result tables and figures from the MA/SMAA model and the MTC/SMAA model, we can see that the outputs from two models do not conflict, but the latter is more discriminative. The differences are described below.

**Estimated effect** After contrasting OR (95% CI) of efficacy from meta-analysis and network meta-analysis, it turns out that almost all the OR from network meta-analysis are larger than those from meta-analysis. The larger OR is, the larger difference between the other drugs and the baseline (Fluoxetine) is. In addition, the OR CI for the measurement are smaller in MTC/SMAA than in MA/SMAA since the former combines more evidence. The interval is 95% CI which describes the uncertainty, so the smaller it is, the lower uncertainty is.

**Rank acceptabilities** The rank acceptabilities in three scenarios from the two models are visualized as column charts shown in Figures 11, 12 and 13. From these figures we can see that in general the results from the MTC/SMAA model are more discriminative than the MA/SMAA ones, especially in the mild depression scenario. For instance, see Figure 12, the rank acceptability for Fluoxetine at the first rank is much higher and it at the fourth rank is much lower in the MTC/SMAA model result compared with the MA/SMAA one. In this case the DMs could almost make the decision of choosing Fluoxetine for mild depression through these results.

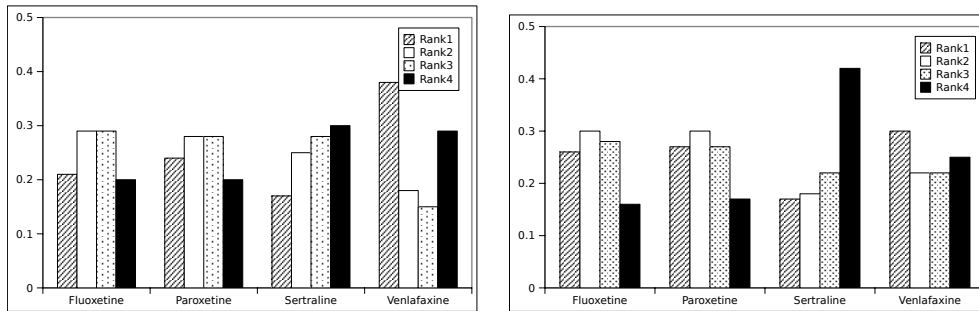


Figure 11: Rank acceptability indices without preference information. Left one is from MA/SMAA and right one from MTC/SMAA.

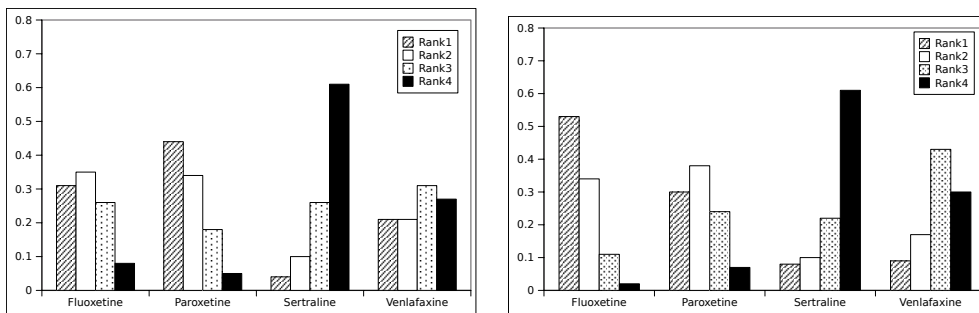


Figure 12: Rank acceptability indices in mild depression scenario. Left one is from MA/SMAA and right one from MTC/SMAA.

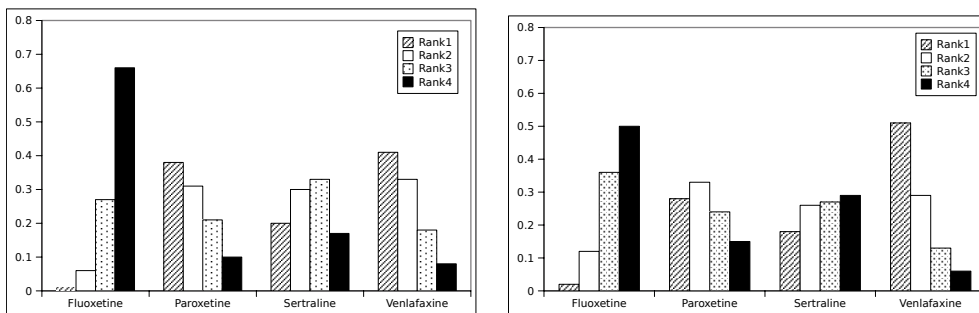


Figure 13: Rank acceptability indices in severe depression scenario. Left one is from MA/SMAA and right one from MTC/SMAA.

By comparing the rank acceptability indices between the two methods, we can see that the performances of the drugs vary in the different scenarios.

Significant variations ( $>0.05$ ) are reported in Table 3. It tells us that in general Fluoxetine’s performance is better in any scenario in MTC/SMAA than in MA/SMAA as its first and second rank acceptabilities increase while the fourth decrease. However, the situation for the other three drugs is mostly the opposite, with the exception of Venlafaxine in the severe depression scenario. From equation 9 we can see that rank acceptability is decided by criteria distribution  $f(\xi)$  and weight distribution  $f(w)$ . According to equations 5 and 6, the weight distribution does not vary in the MA/SMAA model and the MTC/SMAA model as the number of criteria does not change. Therefore the changes of rank acceptabilities come from the differences in criteria distributions. By contrasting Appendix A Table 5 and Appendix C Table 17, it tells us that the SDs of ADRs for Paroxetine, Sertraline and Venlafaxine increase a lot in the MTC/SMAA model compared to the MA/SMAA model. The increased SDs are the result of using a statistic model to estimate the effect instead of just the pooling applied by Hansen et al. [20]. So the criteria distributions with higher standard error contribute to the worse performances in terms of rank acceptability indices for these three drugs. Besides this, another reason for Fluoxetine’s better performance in the MTC/SMAA model might be the original study design. According to Hansen et al. [20], majority of the trials were funded by pharmaceutical companies, and sponsorship was associated with a 5% difference in treatment response (favoring the sponsor’s drug).

Table 3: Variation of rank acceptability indice in MTC/SMAA compared with MA/SMAA and the value is expressed as rank (difference)

drug	no preference	mild depression	severe depression
Fluoxetine	+1(0.06),-4(0.06)	+1(0.23),-3(0.15),-4(0.06)	+2(0.07),-3(0.13),-4(0.21)
Paroxetine	no significant variation	-1(0.15),+2(0.05),+3(0.07)	-1(0.14),+4(0.08)
Sertraline	-2(0.07),-3(0.07),+4(0.12)	no significant variation	-2(0.05),-3(0.1),+4(0.15)
Venlafaxine	-1(0.09),+3(0.07)	-1(0.12),-2(0.06),+3(0.13)	+1(0.11),-3(0.07)

**Central weight vectors** Appendix A Table 7 and Appendix C Table 19 show that there are clear trade-offs among the drugs in both methods. By

contrasting the DM's preferences for scale swings with the central weights, the DMs can decide the preferred drug quickly. For example, if the DMs consider the scale swing of efficacy more important than the scale swing of dizziness, then with the MA/SMAA result they should prefer the BR profile of the three other drugs over Fluoxetine, because it is the only drug for which the central weight of efficacy is considerably lower than the central weight of dizziness [3]. With the MTC/SMAA result they should prefer the BR profile of only Venlafaxine, because it is the only drug for which the central weight of efficacy is considerably higher than the central weight of dizziness.

**Confidence factors** The confidence factors quantify the risk associated with the decision, which means that the low confidence factor shows the uncertainty of making a truly informed decision. Through comparing the confidence factors from two models (Appendix A Table 7 and Appendix C Table 19), we can find that results of Fluoxetine decreased less than of the other three drugs. The possible explanation for this phenomenon can be also the higher SDs of ADRs in the MTC/SMAA model, as the other drugs have less precise ADR measurements than Fluoxetine.

**Overall results** The differences between results from the two models come from the indirect comparisons among Fluoxetine, Paroxetine, Sertraline and Venlafaxine. In Figure 7 we can see that besides the three comparisons analyzed in the three meta-analyses (bold lines), three more comparisons between Paroxetine and Sertraline, between Sertraline and Venlafaxine, and between Paroxetine and Venlafaxine are taken into account by network meta-analysis.

#### 4.3.2 Qualitative comparison

The qualitative comparison is conducted by answering the following questions:

**What is required to implement the method?** To implement the MA/SMAA method, systematic review is applied for data collection, the evaluation criteria are chosen, random-effect meta-analysis is applied for evidence synthesis, and SMAA is used to construct the decision model. To implement MTC/SMAA method, the only difference is that instead of a meta-analysis, a (Bayesian) MTC model has to be constructed. In the Bayesian analysis, the model is constructed (or generated, see [24]) at first and then Markov Chain Monte Carlo simulation is applied to run it.

**What is the difference in the complexity of the methods?** Evaluation on complexity of the methods considers three perspectives:

1. Steps required to implements the methods:

In the MA/SMAA method, data is collected through systematic review of multiple studies, and by statistically combining the data from similar studies the precision of the estimates of treatment effects is improved. In order to combine the results from similar studies, a weighted average is calculated which is an average where the results of some of the studies make a greater contribution to the total than others. Normally trials with higher event rates and smaller CI get a higher weight. Meta-analysis uses statistical techniques to calculate such a weighted average. In the MA/SMAA method, a random effects model is applied to calculate RR, which is an approach to meta-analysis that assumes the true treatment effects in the individual studies are different from each other. As a result, there is a distribution to estimate the treatment effects, and this is used to construct the SMAA model.

In the MTC/SMAA method, the data is also collected through a systematic review but more studies can be included as studies with indirect comparisons are also counted in. All the MTC models are fitted by using Bayesian inference which allows to estimate the relative effects of pairs of drugs. Through Markov Chain Monte Carlo simulation with 50,000 iterations, distributions of numbers are generated in the form of OR (95% CI) which are the relative effects of different drugs. Before constructing the SMAA model, one additional step is needed: construction of inconsistency and consistency models.

The detailed process of this step was shown in Figure 9.

## 2. Model generation:

Compared to the MA/SMAA method, the MTC/SMAA method differs mainly in the MTC part which is more complex than a normal meta-analysis. The MTC is a Bayesian model with a difficult model specification step. In meta-analysis, Escitalopram cannot be included in the analysis as there is no direct comparison between it and Fluoxetine (see Figure 6). In network meta-analysis, all drugs can be included.

## 3. Format of the results:

The results under absolute measurements are slightly different with the results under relative measurements from the MTC/SMAA model (see Appendices B and C), which illustrates that the scale also influences the results besides the parameters in this model. In the scenario with mild depression (see Appendix A Figure 15 and Appendix B Figure 18), efficacy is the last one in the preference list and it is also the only one under the same measurement (Log-OR) in both methods. The differences in scales of ADRs contribute a lot in this scenario, which explains why the results change more than in the other two.

In meta-analysis, the effects can be in RR or OR, but in network meta-analysis, only OR is feasible. From the mathematical point of view, OR of event is reciprocal with OR of non-event. However, with RR there is no such clear relationship between event and non-event, so it requires care in choosing whether to analyze the RR of the event or non-event. Moreover, in consistency model of network meta-analysis, take Figure 10 as an example, if through direct comparison we can get OR between F and P ( $OR_{PF}$ ) and OR between V and F ( $OR_{FV}$ ), then through indirect comparison OR between P and V equals to  $OR_{PF} \times OR_{FV}$ . The equation does not work with RR. Therefore, almost all the MTC models only use OR.

The scales of the MA/SMAA model are in Log-RR and absolute risk that are easier to understand, whereas the scales of the MTC/SMAA model are in Log-OR. The latter has better mathematical properties (symmetric at zero)

and easier to switch “good” and “bad” outcomes<sup>3</sup> than Log-RR and absolute risk, but the numbers are meaningless to the DMs.

**What additional information is required in MTC/SMAA?** The analysis of inconsistency MTC model is required in MTC/SMAA, and the analysis additionally includes determination of inconsistency, source of inconsistency, and elimination of inconsistency.

Understandable scales are extremely important for a decision model to be of any use. Using relative effects (Log-OR) and absolute effects (risk) to measure the criteria give slightly different results. Although the results under relative effects represent the effects more precisely, a relative scale is meaningless to DMs as it cannot represent the results in clinical terms. However, with absolute risk we can get numbers needed to treat (NNT) as  $NNT = 1/(\text{risk difference})$ , where risk difference is the difference between the risk for the baseline and the risk for the other drug. NNT is a common used way to express effect in clinical terms as it gives the number of people who would have to be treated with the experimental intervention to prevent one event compared with the reference. Therefore, the results from the MTC/SMAA model need to be transformed before they can contribute to decision making. To transform the effect into absolute measurement, the risk of baseline must be first assumed and then the absolute risk for each alternative can be calculated with its Log-OR and the assumed baseline risk.

---

<sup>3</sup>Most dichotomised outcomes will be a dichotomy between a good and a bad event [43], and in the RR and OR sense, “good” refers to event and “bad” refers to non-event. There is no direct relationship between the RR of the good outcome and of the bad outcome, but the OR of the good outcome and of the bad outcome is reciprocal with each other.



## 5 Concluding remarks

### 5.1 Conclusion

This thesis introduced a multicriteria decision model with criteria measurements from network meta-analysis (MTC/SMAA) for drug BR analysis which helps pharmacological decision making. A MTC/SMAA model was constructed for four second-generation antidepressants as an example. The results showed clear trade-offs among these antidepressants with multiple criteria model including uncertain criteria measurements, and it is easier for DMs to make decision as the results from this model are shown more discriminative than from the previous model (MA/SMAA) proposed by Tervonen et al. [3]. Compared to the MA/SMAA method, the MTC/SMAA method has three main advantages.

The first advantage is the possibility of taking into account all the available evidence no matter whether the treatments are directly or indirectly related. In the MA/SMAA method, the SMAA model is constructed with criteria measurements from pairwise treatment comparisons which means that only the evidence from direct comparisons can be included. With the MTC/SMAA method, the SMAA model is built with criteria measurements from mixed treatment comparisons which estimate the treatment effect by synthesizing all the available evidence.

The second advantage is the possibility to decide upon the “best” treatment in a class without a common comparator. In the MA/SMAA method, one drug is chosen to be the common comparator and all the other drugs included in the analysis must have a direct relationship with it, otherwise the MA/SMAA model cannot be applied. If the common comparator is changed, the results might change. With the MTC/SMAA method, it does not matter which drug is the common comparator as it doesn’t change the results.

The third advantage is the possibility to detect evidence inconsistency when the studies providing evidence have different designs and consequently improve the transparency of decision making process.

However, according to the qualitative analysis in this thesis, the disadvantage is its considerable complexity of implementation compared with the MA/SMAA method.

## **5.2 Future work**

In this thesis, the constructed model is specific to the antidepressants and has to be done in two steps, by first constructing the MTC models and then the SMAA model, separately and manually, which requires specific knowledge from DMs and is not easy for them to apply the models. Especially in the MTC model generation, the analysis of inconsistency is complicated and beyonds the knowledge of normal DMs, and the criteria measurements need to be transformed into understandable format. The future research should focus on the follows.

1. Simplifying the analysis of inconsistency.
2. Automation transformation of the criteria measurements into understandable format.
3. Applicability of the MTC/SMAA method to other therapeutic areas.
4. Integration of the MTC model construction, the SMAA model construction and the BR analysis into a single automated decision support process (undergoing research, see: [www.drugis.org](http://www.drugis.org)).

## References

- [1] N. Victor, J. Hasford, Risk-benefit analyses of drugs: fundamental considerations and requirements from the point of view of the biometrician. problems in the assessment of the combination of trimethoprim with sulfamethoxazole, *Infection* 15 (1987) 236–240.
- [2] D. Ouellet, Benefit-risk assessment: the use of clinical utility index, *Expert Opinion on Drug Safety* 9(2) (2010) 289–300.
- [3] T. Tervonen, H. L. Hillege, E. Buskens, D. Postmus, A state-of-the-art multi-criteria model for drug benefit-risk analysis, SOM Research Report 10001, Faculty of Economics and Business, University of Groningen, The Netherlands (2010).
- [4] L. D. Lynd, B. J. O'Brien, Advances in risk-benefit evaluation using probabilistic simulation methods: an application to the prophylaxis of deep vein thrombosis, *Journal of Clinical Epidemiology* 57(8) (2004) 795–803.
- [5] A. S. Elstein, On the origins and development of evidence-based medicine and medical decision making., *Inflammation Research* 53(2) (2004) 184–189.
- [6] D. Gough, Weight of evidence: a framework for the appraisal of the quality and relevance of evidence, *Research Papers in Education* 22(2) (2007) 213–228.
- [7] T. Lumley, Network meta-analysis for indirect treatment comparisons, *Statistics in Medicine* 21 (16) (2002) 2313–2324.
- [8] G. Lu, A. E. Ades, Assessing evidence inconsistency in mixed treatment comparisons, *Journal of the American Statistical Association* 101 (474) (2006) 447–459.
- [9] G. Salanti, J. P. T. Higgins, A. E. Ades, J. P. A. Ioannidis, Evaluation of networks of randomized trials, *Statistical Methods in Medical Research* 17 (3) (2008) 279–301.

- [10] V. Belton, T. J. Stewart, *Multiple Criteria Decision Analysis - An Integrated Approach*, Kluwer Academic Publishers, Dordrecht, 2002.
- [11] R. Keeney, H. Raiffa, *Decisions with multiple objectives: preferences and value tradeoffs*, Wiley, New York, 1976.
- [12] F. Mussen, S. Salek, S. Walker, A quantitative approach to benefit-risk assessment of medicines – part 1: The development of a new model using multi-criteria decision analysis, *Pharmacoepidemiology and Drug Safety* 16 (2007) S12–S15.
- [13] J. C. Felli, R. A. Noel, P. A. Cavazzoni, A multiattribute model for evaluating the benefit-risk profiles of treatment alternatives, *Medical Decision Making* 29 (1) (2009) 104–115.
- [14] R. Lahdelma, J. Hokkanen, P. Salminen, SMAA - stochastic multiobjective acceptability analysis, *European Journal of Operational Research* 106 (1) (1998) 137–143.
- [15] J. Hokkanen, R. Lahdelma, P. Salminen, A multiple criteria decision model for analyzing and choosing among different development patterns for the Helsinki cargo harbor, *Socio-Economic Planning Sciences* 33 (1999) 1–23.
- [16] T. Tervonen, H. Hakonen, R. Lahdelma, Elevator planning with stochastic multicriteria acceptability analysis, *Omega* 36(3) (2008) 352–362.
- [17] J. Kangas, J. Hokkanen, A. Kangas, R. Lahdelma, P. Salminen, Applying stochastic multicriteria acceptability analysis to forest ecosystem management with both cardinal and ordinal criteria, *Forest Science* 49 (6) (2003) 928–937.
- [18] A. Menou, A. Benallou, R. Lahdelma, P. Salminen, Decision support for centralizing cargo at a moroccan airport hub using stochastic multicriteria acceptability analysis, *European Journal of Operational Research*.

- [19] S. Makkonen, R. Lahdelma, A. M. Asel, A. Jokinen, Multicriteria decision support in the liberated energy market, *Journal of Multi-Criteria Decision Analysis* 12(1) (2003) 27–42.
- [20] R. A. Hansen, G. Gartlehner, K. N. Lohr, B. N. Gaynes, T. Carey, Efficacy and safety of second-generation antidepressants in the treatment of major depressive disorder, *Annals of Internal Medicine* 143 (6) (2005) 415–426.
- [21] A. J. Sutton, J. P. T. Higgins, Recent developments in meta-analysis, *Statistics in Medicine* 18 (3) (2008) 625–650.
- [22] P. Karl, Report on certain enteric fever inoculation statistics., *British Medical Journal* 3 (1904) 1243–1246.
- [23] A. Sutton, A. E. Ades, N. Cooper, K. Abrams, Use of indirect and mixed treatment comparisons for technology assessment, *Pharmacoeconomics* 26 (9) (2008) 753–767.
- [24] G. van Valkenhoef, T. Tervonen, B. de Brock, H. L. Hillege, Algorithmic parametrization of mixed treatment comparisons, manuscript.
- [25] R. Lahdelma, K. Miettinen, P. Salminen, Ordinal criteria in stochastic multicriteria acceptability analysis (SMAA), *European Journal of Operational Research* 147 (1) (2003) 117–127.
- [26] T. Tervonen, New directions in stochastic multicriteria acceptability analysis, Ph.D. thesis, University of Turku (2007).
- [27] R. Lahdelma, P. Salminen, SMAA-2: Stochastic multicriteria acceptability analysis for group decision making, *Operations Research* 49 (3) (2001) 444–454.
- [28] R. Lahdelma, P. Salminen, Pseudo-criteria versus linear utility function in stochastic multi-criteria acceptability analysis, *European Journal of Operational Research* 141 (2) (2002) 454–469.

- [29] T. Tervonen, J. Figueira, R. Lahdelma, P. Salminen, An inverse approach for ELECTRE III, Tech. rep., Research Report 20/2004 of The Institute of Systems Engineering and Computers (INESC-Coimbra), Coimbra, Portugal (2004).
- [30] R. Lahdelma, P. Salminen, Stochastic multicriteria acceptability analysis using the data envelopment model, *European Journal of Operational Research* 170 (1) (2006) 241–252.
- [31] R. Lahdelma, P. Salminen., Prospect theory and stochastic multicriteria acceptability analysis (SMAA), *Omega* 37(5) (2009) 961–971.
- [32] I. Durbach, A simulation-based test of stochastic multicriteria acceptability analysis using achievement functions, *European Journal of Operational Research* 170 (3) (2006) 923–934.
- [33] T. Tervonen, J. R. Figueira, R. Lahdelma, J. Almeida Dias, P. Salminen, A stochastic method for robustness analysis in sorting problems, *European Journal of Operational Research* 192 (1) (2009) 236–242.
- [34] T. Tervonen, R. Lahdelma, Implementing stochastic multicriteria acceptability analysis, *European Journal of Operational Research* 178 (2) (2007) 500–513.
- [35] N. K. D. . Y. S. Lincoln, *The Sage Handbook of Qualitative Research* (3rd ed.), Thousand Oaks, CA: Sage, 2005.
- [36] R. K. Yin, *Case Study Research: Design and Methods. Fourth Edition.*, SAGE Publications., 2009.
- [37] B. Flyvbjerg, Five misunderstandings about case study research, *Qualitative Inquiry* 12 (2006) 219–245.
- [38] S. W. . A. Romney, *Systematic Data Collection (Qualitative Research Methods Series 10)*, SAGE Publications, 1988.
- [39] P. Martyn. JAGS: Just Another Gibbs Sampler. [on-line]. Accessed on May 17, 2010 at URL: <http://www-fis.iarc.fr/martyn/software/jags>.

- [40] R Development Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0 (2008).  
URL <http://www.R-project.org>
- [41] T. Tervonen, JSMAA: an open source software for SMAA computations, in: C. Henggeler Antunes, D. Rios Insua, L.C. Dias (eds.), Proceedings of the 25th Mini EURO Conference on Uncertainty and Robustness in Planning and Decision Making (URPDM2010), April 15-17, 2010, Coimbra, Portugal, 2010.
- [42] L. V. Hedges, J. L. Vevea, Fixed- and random-effects models in meta-analysis, *Psychological Methods* 3 (4) (1998) 486–504.
- [43] The Cochrane Collaboration. [on-line]. Accessed on May 20, 2010 at URL:<http://www.cochrane-net.org/openlearning/index.htm>.

# Appendices

## Appendix A

Table 4: MA/SMAA, criteria characteristics

Name	Measurement unit	Preference direction	Scale range
Efficacy	Relative value compared with Fluoxetine	Ascending	[0.98,1.23]
Diarrhea	Absolute%	Descending	[1,20.6]
Dizziness	Absolute%	Descending	[4.4,24.4]
Headache	Absolute%	Descending	[8,31.3]
Insomnia	Absolute%	Descending	[3.4,21.3]
Nausea	Absolute%	Descending	[11.1,34]

Table 5: MA/SMAA, criteria measurements. The values are given as mean  $\pm$  standard deviation. The measurement units are as presented in Table 4.

Drug	Ln(Efficacy)	Diarrhea	Dizziness	Headache	Insomnia	Nausea
Fluoxetine	0 $\pm$ 0	11.7 $\pm$ 2.5	7.2 $\pm$ 1.45	16.6 $\pm$ 3.27	13.7 $\pm$ 1.89	18.6 $\pm$ 1.79
Paroxetine	0.086 $\pm$ 0.056	9.2 $\pm$ 1.86	10.6 $\pm$ 1.58	21.2 $\pm$ 5.15	14.3 $\pm$ 2.93	18.3 $\pm$ 3.7
Sertraline	0.095 $\pm$ 0.044	15.4 $\pm$ 2.56	7.5 $\pm$ 1.48	20.2 $\pm$ 3.78	15 $\pm$ 3.21	19.5 $\pm$ 2.6
Venlafaxine	0.113 $\pm$ 0.048	5.5 $\pm$ 2.32	15.7 $\pm$ 4.44	12.8 $\pm$ 2.45	11.2 $\pm$ 3.98	31 $\pm$ 1.68

Table 6: MA/SMAA, rank acceptability indices from the analysis without preference information

Drug	Rank1	Rank2	Rank3	Rank4
Fluoxetine	0.20	0.28	0.30	0.22
Paroxetine	0.25	0.29	0.27	0.19
Sertraline	0.17	0.25	0.29	0.30
Venlafaxine	0.39	0.18	0.15	0.29



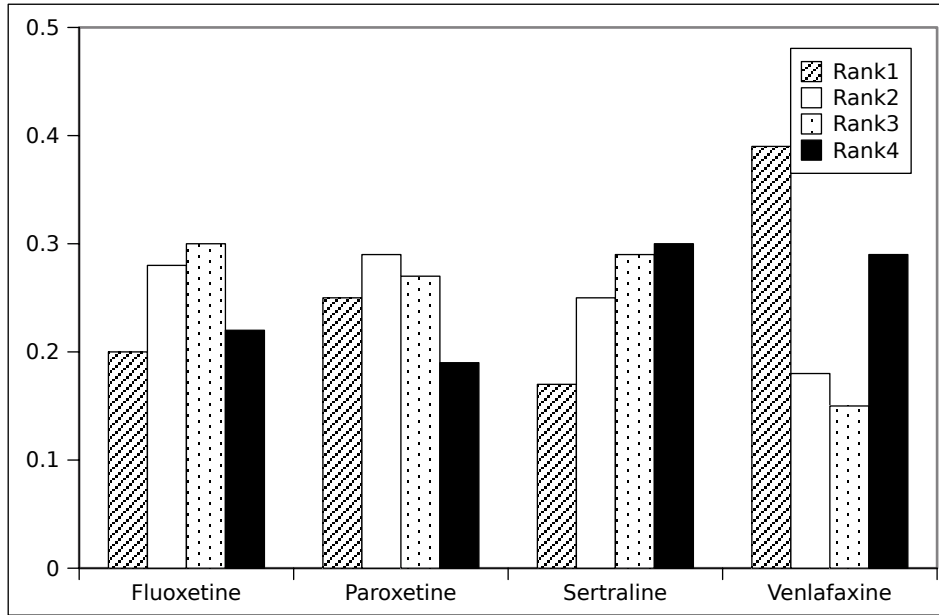


Figure 14: MA/SMAA, rank acceptability indices for the model without preference information.

Table 7: MA/SMAA, central weights and corresponding confidence factors from the analysis without preference information

Drug	Confidence factor	Central weight					
		Efficacy	Diarrhea	Dizziness	Headache	Insomnia	Nausea
Fluoxetine	0.48	0.08	0.14	0.23	0.18	0.16	0.22
Paroxetine	0.45	0.18	0.17	0.15	0.13	0.15	0.22
Sertaline	0.34	0.21	0.10	0.22	0.13	0.15	0.20
Venlafaxine	0.74	0.18	0.21	0.12	0.21	0.19	0.09

Table 8: MA/SMAA, rank acceptability indices from the scenario of mild depression

Drug	Rank1	Rank2	Rank3	Rank4
Fluoxetine	0.30	0.35	0.26	0.08
Paroxetine	0.45	0.33	0.17	0.05
Sertraline	0.04	0.10	0.26	0.60
Venlafaxine	0.21	0.23	0.30	0.26

Table 9: MA/SMAA, rank acceptability indices from the scenario of severe depression

Drug	Rank1	Rank2	Rank3	Rank4
Fluoxetine	0.01	0.05	0.23	0.71
Paroxetine	0.42	0.31	0.20	0.07
Sertraline	0.18	0.31	0.37	0.14
Venlafaxine	0.40	0.32	0.20	0.08

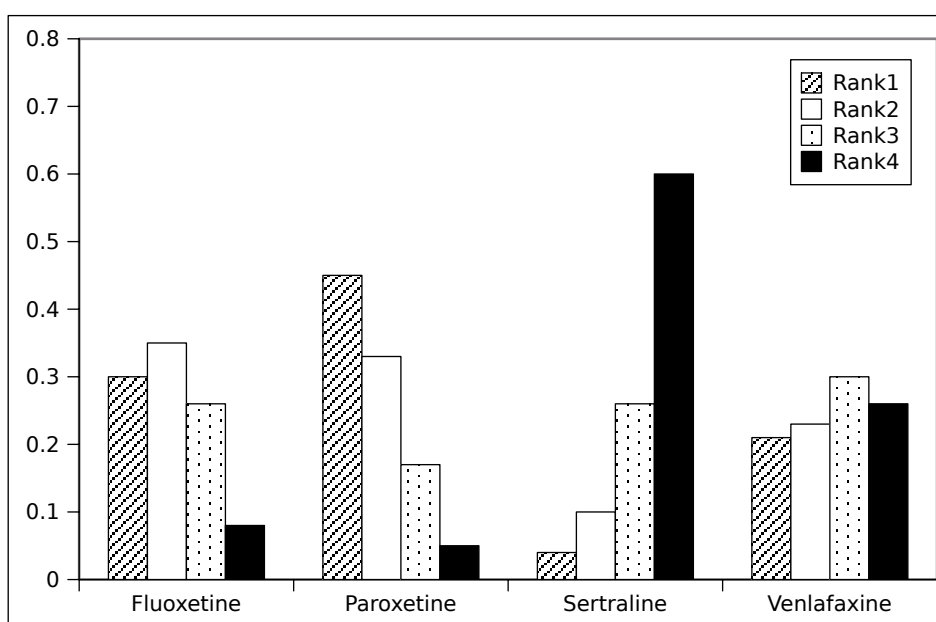


Figure 15: MA/SMAA, rank acceptability indices from the scenario of mild depression.

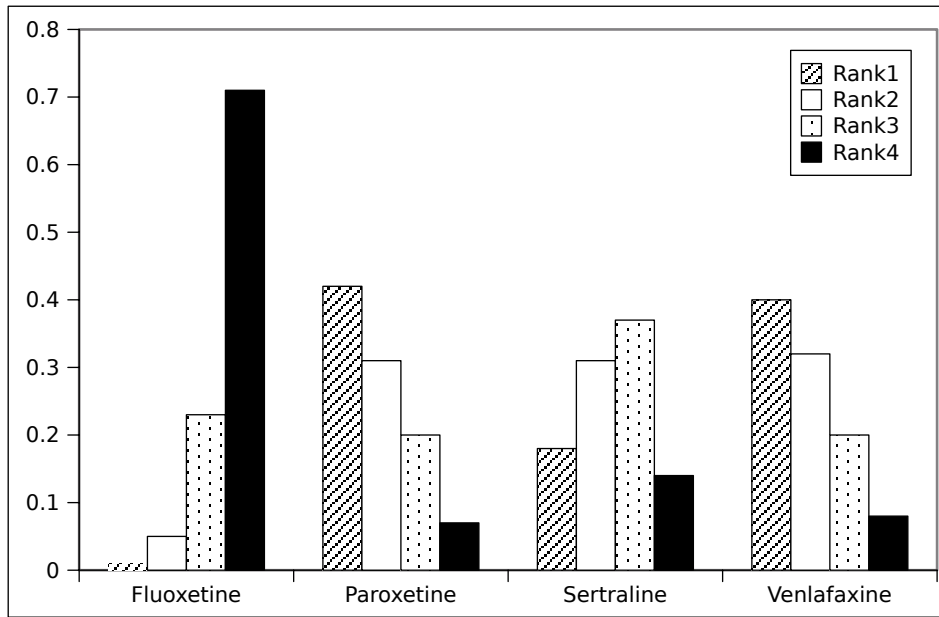


Figure 16: MA/SMAA, rank acceptability indices from the scenario of severe depression.

## Appendix B

Table 10: MTC/SMAA, criteria characteristics

Name	Measurement unit	Preference direction	Scale range
Efficacy	relative value compared with Fluoxetine	Ascending	[0.90-1.76]
Diarrhea	relative value compared with Fluoxetine	Descending	[0.11-10.03]
Dizziness	relative value compared with Fluoxetine	Descending	[0.33-4.98]
Headache	relative value compared with Fluoxetine	Descending	[0.36-2.38]
Insomnia	relative value compared with Fluoxetine	Descending	[0.34-6.66]
Nausea	relative value compared with Fluoxetine	Descending	[0.77-3.03]

Table 11: MTC/SMAA, criteria measurements. The values are given as mean  $\pm$  standard deviation. The measurement units are as presented in Table 10.

Drug	Efficacy	Diarrhea	Dizziness	Headache	Insomnia	Nausea
Fluoxetine	0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0	0 $\pm$ 0
Paroxetine	0.169 $\pm$ 0.138	-0.75 $\pm$ 0.6	0.512 $\pm$ 0.382	-0.137 $\pm$ 0.28	0.213 $\pm$ 0.595	0.29 $\pm$ 0.25
Sertraline	0.242 $\pm$ 0.124	1.07 $\pm$ 0.63	-0.336 $\pm$ 0.376	0.279 $\pm$ 0.3	0.594 $\pm$ 0.664	0.29 $\pm$ 0.28
Venlafaxine	0.348 $\pm$ 0.111	-0.759 $\pm$ 0.72	1.06 $\pm$ 0.279	-0.414 $\pm$ 0.31	0.143 $\pm$ 0.62	0.62 $\pm$ 0.25

Table 12: MTC/SMAA, rank acceptability indices from the analysis without preference information

Drug	Rank1	Rank2	Rank3	Rank4
Fluoxetine	0.33	0.31	0.25	0.11
Paroxetine	0.23	0.33	0.28	0.15
Sertraline	0.15	0.19	0.23	0.43
Venlafaxine	0.29	0.17	0.24	0.30

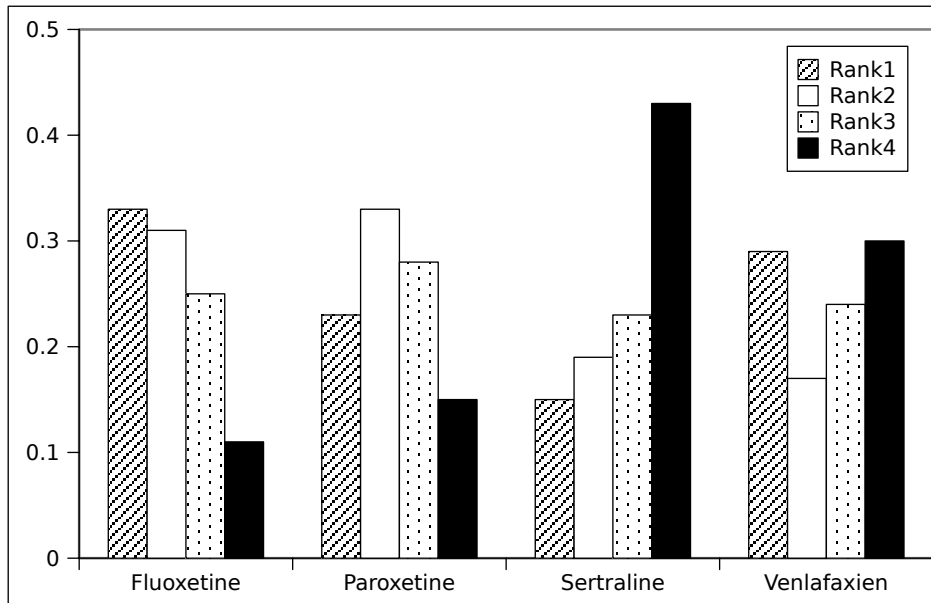


Figure 17: MTC/SMAA, rank acceptability indices for the model without preference information.

Table 13: MTC/SMAA, central weight and corresponding confidence factors from the analysis without preference information

Drug	Confidence factor	Central weight					
		Efficacy	Diarrhea	Dizziness	Headache	Insomnia	Nausea
Fluoxetine	0.76	0.08	0.16	0.20	0.14	0.18	0.24
Paroxetine	0.33	0.16	0.19	0.14	0.18	0.16	0.15
Sertraline	0.30	0.23	0.11	0.24	0.12	0.14	0.16
Venlafaxine	0.61	0.24	0.18	0.10	0.21	0.16	0.11

Table 14: MTC/SMAA, rank acceptability indices from the scenario of mild depression

Drug	Rank1	Rank2	Rank3	Rank4
Fluoxetine	0.77	0.21	0.01	0
Paroxetine	0.19	0.55	0.21	0.06
Sertraline	0.02	0.12	0.32	0.53
Venlafaxine	0.02	0.12	0.46	0.41

Table 15: MTC/SMAA, rank acceptability indices from the scenario of severe depression

Drug	Rank1	Rank2	Rank3	Rank4
Fluoxetine	0.03	0.18	0.40	0.39
Paroxetine	0.25	0.31	0.24	0.19
Sertraline	0.22	0.25	0.22	0.31
Venlafaxine	0.49	0.26	0.14	0.11

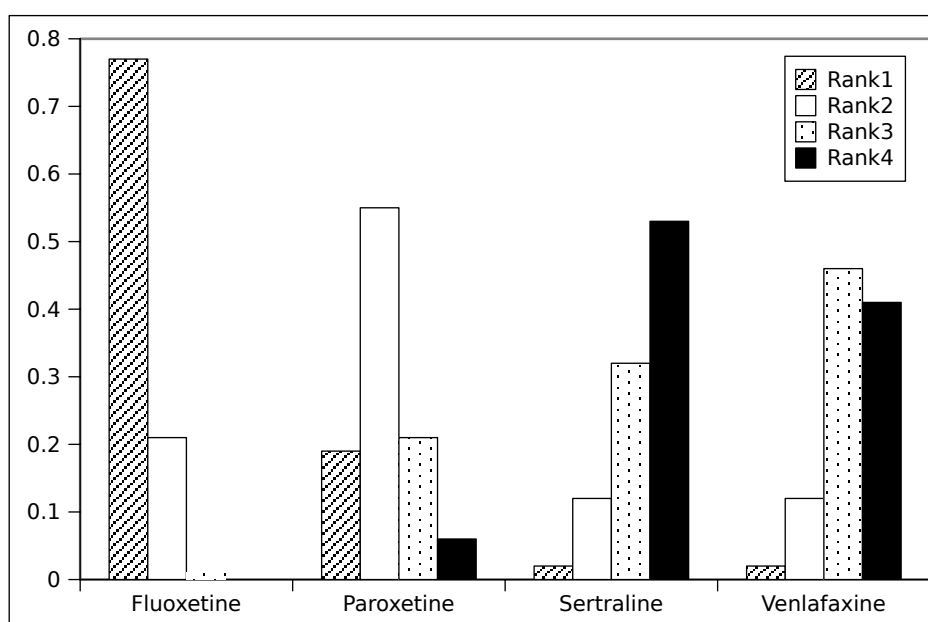


Figure 18: MTC/SMAA, rank acceptability indices from the scenario of mild depression.

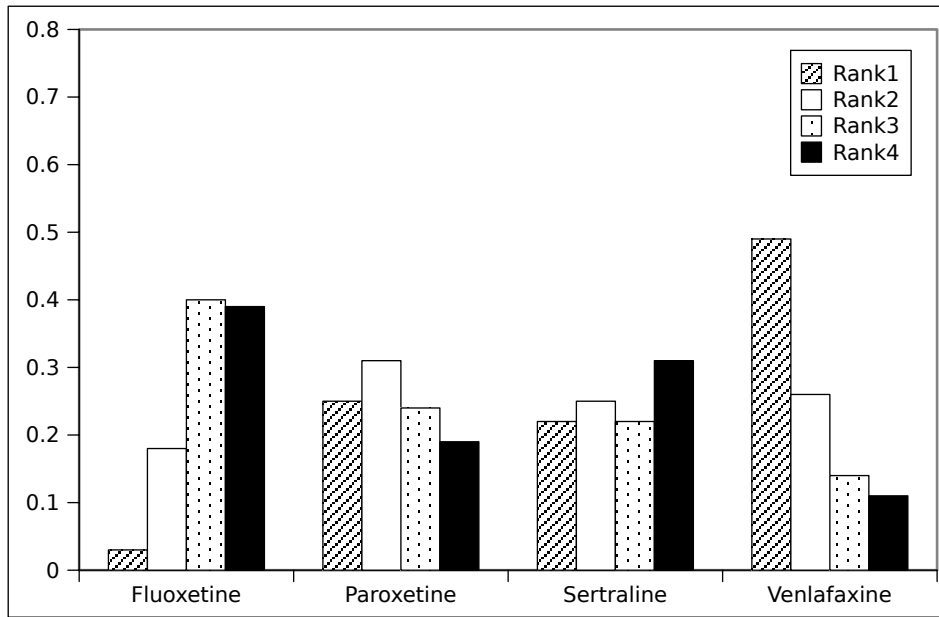


Figure 19: MTC/SMAA, rank acceptability indices from the scenario of severe depression.

## Appendix C

Table 16: MTC/SMAA in absolute measurement, criteria characteristics

Name	Measurement unit	Preference direction	Scale range
Efficacy	Relative value compared with Fluoxetine	Ascending	[0.9,1.76]
Diarrhea	Absolute%	Descending	[-2.24,46.95]
Dizziness	Absolute%	Descending	[1.69,29.02]
Headache	Absolute%	Descending	[6.34,33.98]
Insomnia	Absolute%	Descending	[-0.27,44.94]
Nausea	Absolute%	Descending	[13.81,40.89]

Table 17: MTC/SMAA in absolute measurement, criteria measurements. The values are given as mean  $\pm$  standard deviation. The measurement units are as presented in Table 16.

Drug	Efficacy	Diarrhea	Dizziness	Headache	Insomnia	Nausea
Fluoxetine	0 $\pm$ 0	7.8 $\pm$ 3	6.7 $\pm$ 2	16.8 $\pm$ 4.4	10.5 $\pm$ 4.2	17.9 $\pm$ 3.4
Paroxetine	0.169 $\pm$ 0.138	4.5 $\pm$ 3.4	11.1 $\pm$ 5.2	15.2 $\pm$ 5.4	13.9 $\pm$ 9	22.8 $\pm$ 6
Sertraline	0.242 $\pm$ 0.124	21.3 $\pm$ 12.1	5 $\pm$ 2.4	21.3 $\pm$ 7.2	19.2 $\pm$ 12	22.8 $\pm$ 6.4
Venlafaxine	0.348 $\pm$ 0.111	4.8 $\pm$ 4.1	17.2 $\pm$ 6.1	12.1 $\pm$ 4.8	13.3 $\pm$ 8.7	28.8 $\pm$ 6.8

Table 18: MTC/SMAA in absolute measurement, rank acceptability indices from the analysis without preference information

Drug	Rank1	Rank2	Rank3	Rank4
Fluoxetine	0.26	0.3	0.28	0.16
Paroxetine	0.27	0.3	0.27	0.17
Sertraline	0.17	0.18	0.22	0.42
Venlafaxine	0.3	0.22	0.22	0.25



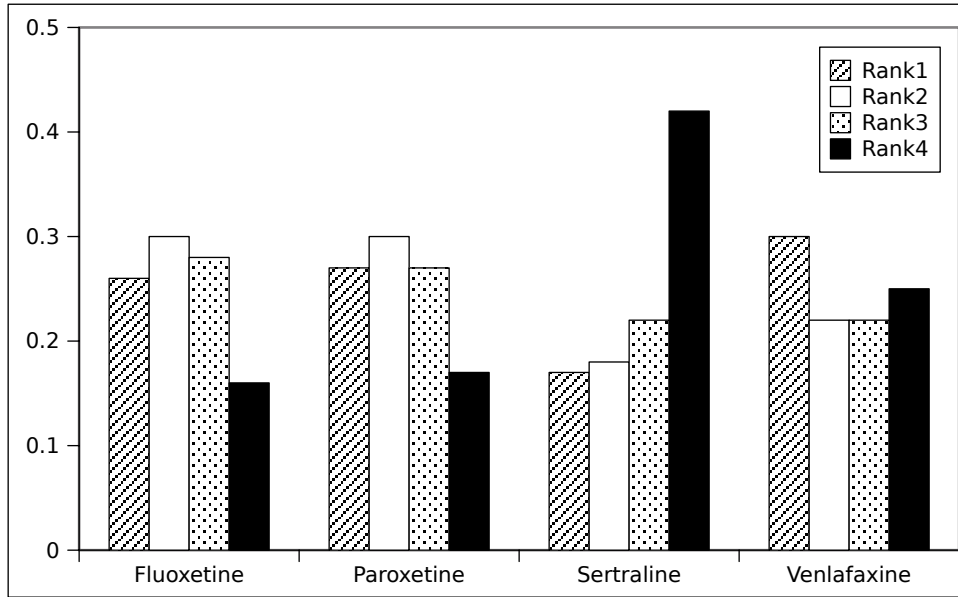


Figure 20: MTC/SMAA in absolute measurement, rank acceptability indices for the model without preference information.

Table 19: MTC/SMAA in absolute measurement, central weight and corresponding confidence factors from the analysis without preference information

Drug	Confidence factor	Central weight					
		Efficacy	Diarrhea	Dizziness	Headache	Insomnia	Nausea
Fluoxetine	0.49	0.09	0.16	0.2	0.15	0.18	0.22
Paroxetine	0.34	0.16	0.19	0.16	0.17	0.17	0.16
Sertaline	0.23	0.19	0.12	0.23	0.14	0.15	0.17
Venlafaxine	0.52	0.23	0.18	0.11	0.2	0.17	0.12

## Appendix D

Burke WJ, Gergel I, Bose A. Fixed-dose trial of the single isomer SSRI escitalopram in depressed outpatients. *J Clin Psychiatry*. 2002;63:331-6. [PMID: 12000207]

Lepola UM, Loft H, Reines EH. Escitalopram (10-20 mg/day) is effective and well tolerated in a placebo-controlled study in depression in primary care. *Int Clin Psychopharmacol*. 2003;18:211-7. [PMID: 12817155]

Patris M, Bouchard JM, Bougerol T, Charbonnier JF, Chevalier JF, Clerc G,

Table 20: MTC/SMAA in absolute measurement, rank acceptability indices from the scenario of mild depression

Drug	Rank1	Rank2	Rank3	Rank4
Fluoxetine	0.53	0.34	0.11	0.02
Paroxetine	0.3	0.38	0.24	0.07
Sertraline	0.08	0.1	0.22	0.61
Venlafaxine	0.09	0.17	0.43	0.3

Table 21: MTC/SMAA in absolute measurements, rank acceptability indices from the scenario of severe depression

Drug	Rank1	Rank2	Rank3	Rank4
Fluoxetine	0.02	0.12	0.36	0.5
Paroxetine	0.28	0.33	0.24	0.15
Sertraline	0.18	0.26	0.27	0.29
Venlafaxine	0.51	0.29	0.13	0.06

et al. Citalopram versus fluoxetine: a double-blind, controlled, multicentre, phase III trial in patients with unipolar major depression treated in general practice. *Int Clin Psychopharmacol.* 1996;11:129-36. [PMID: 8803650]

Ekselius L, von Knorring L, Eberhard G. A double-blind multicenter trial comparing sertraline and citalopram in patients with major depression treated in general practice. *Int Clin Psychopharmacol.* 1997;12:323-31. [PMID: 9547134]

Dalery J, Honig A. Fluvoxamine versus fluoxetine in major depressive episode: a double-blind randomised comparison. *Hum Psychopharmacol.* 2003;18:379-84. [PMID: 12858325]

Rapaport M, Coccaro E, Sheline Y, Perse T, Holland P, Fabre L, et al. A comparison of fluvoxamine and fluoxetine in the treatment of major depression. *J Clin Psychopharmacol.* 1996;16:373-8. [PMID: 8889909]

Chouinard G, Saxena B, Belanger MC, Ravindran A, Bakish D, Beauclair L, et al. A Canadian multicenter, double-blind study of paroxetine and fluoxetine in major depressive disorder. *J Affect Disord.* 1999;54:39-48. [PMID: 10403145]

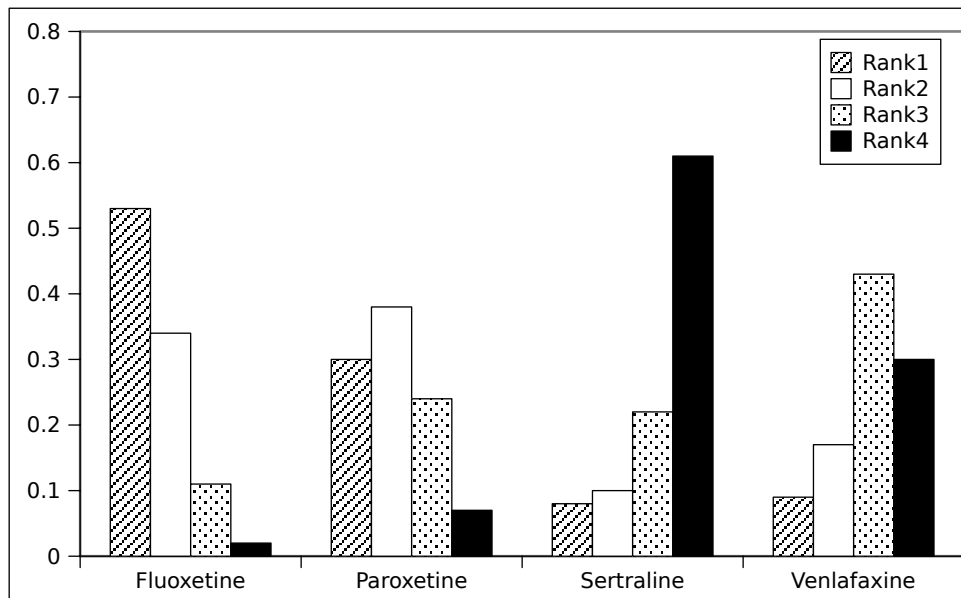


Figure 21: MTC/SMAA in absolute measurement, rank acceptability indices from the scenario of mild depression.

De Wilde J, Spiers R, Mertens C, Bartholome F, Schotte G, Leyman S. A double-blind, comparative, multicentre study comparing paroxetine with fluoxetine in depressed patients. *Acta Psychiatr Scand.* 1993;87:141-5. [PMID: 8447241]

Schone W, Ludwig M. A double-blind study of paroxetine compared with fluoxetine in geriatric patients with major depression. *J Clin Psychopharmacol.* 1993;13(6 Suppl 2):34S-39S. [PMID: 8106654]

Fava M, Amsterdam JD, Deltito JA, Salzman C, Schwaller M, Dunner DL. A double-blind study of paroxetine, fluoxetine, and placebo in outpatients with major depression. *Ann Clin Psychiatry.* 1998;10:145-50. [PMID: 9988054]

Fava M, Hoog SL, Judge RA, Kopp JB, Nilsson ME, Gonzales JS. Acute efficacy of fluoxetine versus sertraline and paroxetine in major depressive disorder including effects of baseline insomnia. *J Clin Psychopharmacol.* 2002;22:137-47. [PMID: 11910258]

Gagiano CA. A double blind comparison of paroxetine and fluoxetine in patients with major depression. *Br J Clin Res.* 1993;4:145-52.

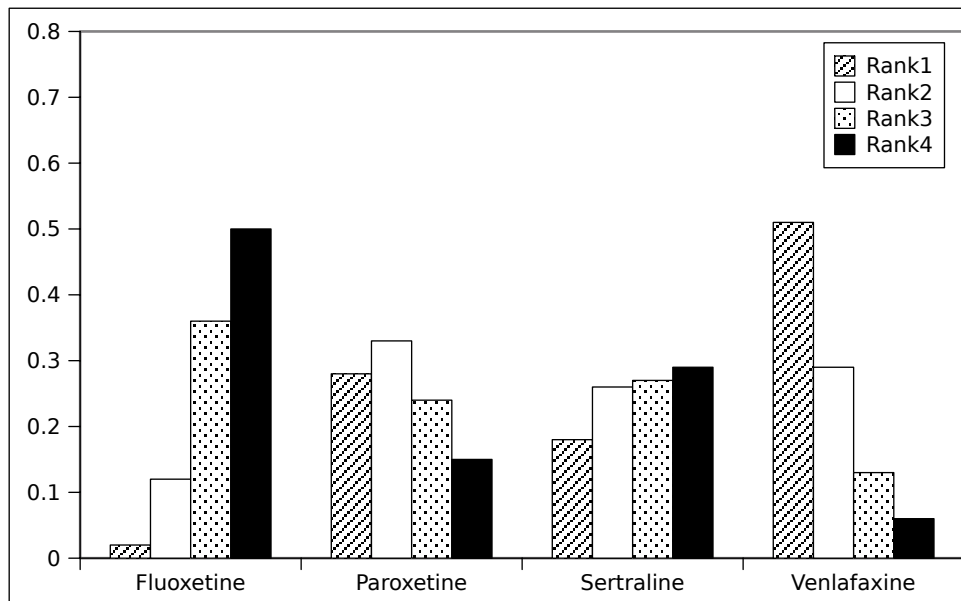


Figure 22: MTC/SMAA in absolute measurement, rank acceptability indices from the scenario of severe depression.

Bennie EH, Mullin JM, Martindale JJ. A double-blind multicenter trial comparing sertraline and fluoxetine in outpatients with major depression. *J Clin Psychiatry*. 1995;56:229-37. [PMID: 7775364]

Boyer P, Danion JM, Bisserbe JC, Hotton JM, Troy S. Clinical and economic comparison of sertraline and fluoxetine in the treatment of depression. A 6-month double-blind study in a primary-care setting in France. *Pharmacoeconomics*. 1998;13(1 Pt 2):157-69. [PMID: 10184835]

Sechter D, Troy S, Paternetti S, Boyer P. A double-blind comparison of sertraline and fluoxetine in the treatment of major depressive episode in outpatients. *Eur Psychiatry*. 1999;14:41-8. [PMID: 10572324]

Newhouse PA, Krishnan KR, Doraiswamy PM, Richter EM, Batzar ED, Clary CM. A double-blind comparison of sertraline and fluoxetine in depressed elderly outpatients. *J Clin Psychiatry*. 2000;61:559-68. [PMID: 10982198]

Kiev A, Feiger A. A double-blind comparison of fluvoxamine and paroxetine in the treatment of depressed outpatients. *J Clin Psychiatry*. 1997;58:146-52.

[PMID: 9164424]

Aberg-Wistedt A, Agren H, Ekselius L, Bengtsson F, Akerblad AC. Sertraline versus paroxetine in major depression: clinical outcome after six months of continuous therapy. *J Clin Psychopharmacol.* 2000;20:645-52. [PMID: 11106136]

Nemeroff CB, Ninan PT, Ballenger J, Lydiard RB, Feighner J, Patterson WM, et al. Double-blind multicenter comparison of fluvoxamine versus sertraline in the treatment of depressed outpatients. *Depression.* 1995;3:163-9.

Franchini L, Gasperini M, Perez J, Smeraldi E, Zanardi R. A double-blind study of long-term treatment with sertraline or fluvoxamine for prevention of highly recurrent unipolar depression. *J Clin Psychiatry.* 1997;58:104-7. [PMID: 9108811]

Bielski RJ, Ventura D, Chang CC. A double-blind comparison of escitalopram and venlafaxine extended release in the treatment of major depressive disorder. *J Clin Psychiatry.* 2004;65:1190-6. [PMID: 15367045]

Montgomery SA, Huusom AK, Bothmer J. A randomised study comparing escitalopram with venlafaxine XR in primary care patients with major depressive disorder. *Neuropsychobiology.* 2004;50:57-64. [PMID: 15179022]

Feighner JP, Gardner EA, Johnston JA, Batey SR, Khayrallah MA, Ascher JA, et al. Double-blind comparison of bupropion and fluoxetine in depressed outpatients. *J Clin Psychiatry.* 1991;52:329-35. [PMID: 1907963]

Coleman CC, King BR, Bolden-Watson C, Book MJ, Segraves RT, Richard N, et al. A placebo-controlled comparison of the effects on sexual functioning of bupropion sustained release and fluoxetine. *Clin Ther.* 2001;23:1040-58. [PMID: 11519769]

Hong CJ, Hu WH, Chen CC, Hsiao CC, Tsai SJ, Ruwe FJ. A doubleblind, randomized, group-comparative study of the tolerability and efficacy of 6 weeks' treatment with mirtazapine or fluoxetine in depressed Chinese patients. *J Clin Psychiatry.* 2003;64:921-6. [PMID: 12927007]

Goldstein DJ, Mallinckrodt C, Lu Y, Demitrack MA. Duloxetine in the treat-

ment of major depressive disorder: a double-blind clinical trial. *J Clin Psychiatry*. 2002;63:225-31. [PMID: 11926722]

Alves C, Cachola I, Brandao J. Efficacy and tolerability of venlafaxine and fluoxetine in outpatients with major depression. *Primary Care Psychiatry*. 1999; 5:57-63.

Tylee A, Beaumont G, Bowden MW, Reynolds A. A double-blind, randomized, 12-week comparison study of the safety and efficacy of venlafaxine and fluoxetine in moderate to severe depression in general practice. *Primary Care Psychiatry*. 1997;3:51-8.

Dierick M, Ravizza L, Realini R, Martin A. A double-blind comparison of venlafaxine and fluoxetine for treatment of major depression in outpatients. *Prog Neuropsychopharmacol Biol Psychiatry*. 1996;20:57-71. [PMID: 8861177]

De Nayer A, Geerts S, Ruelens L, Schittecatte M, De Bleeker E, Van Eeckhoutte I, et al. Venlafaxine compared with fluoxetine in outpatients with depression and concomitant anxiety. *Int J Neuropsychopharmacol*. 2002;5:115-20. [PMID: 12135535]

Rudolph RL, Feiger AD. A double-blind, randomized, placebo-controlled trial of once-daily venlafaxine extended release (XR) and fluoxetine for the treatment of depression. *J Affect Disord*. 1999;56:171-81. [PMID: 10701474]

Silverstone PH, Ravindran A. Once-daily venlafaxine extended release (XR) compared with fluoxetine in outpatients with depression and anxiety. Venlafaxine XR 360 Study Group. *J Clin Psychiatry*. 1999;60:22-8. [PMID: 10074873]

Weihs KL, Settle EC Jr, Batey SR, Houser TL, Donahue RM, Ascher JA. Bupropion sustained release versus paroxetine for the treatment of depression in the elderly. *J Clin Psychiatry*. 2000;61:196-202. [PMID: 10817105]

Doraiswamy PM, Khan ZM, Donahue RM, Richard NE. Quality of life in geriatric depression: a comparison of remitters, partial responders, and non-responders. *Am J Geriatr Psychiatry*. 2001;9:423-8. [PMID: 11739069]

Detke MJ, Wiltse CG, Mallinckrodt CH, McNamara RK, Demitrack MA, Bitter I. Duloxetine in the acute and long-term treatment of major depressive

disorder: a placebo- and paroxetine-controlled trial. *Eur Neuropsychopharmacol.* 2004;14:457-70. [PMID: 15589385]

Schatzberg AF, Kremer C, Rodrigues HE, Murphy GM Jr. Double-blind, randomized comparison of mirtazapine and paroxetine in elderly depressed patients. Mirtazapine vs. Paroxetine Study Group. *Am J Geriatr Psychiatry.* 2002; 10:541-50. [PMID: 12213688]

Benkert O, Szegedi A, Kohnen R. Mirtazapine compared with paroxetine in major depression. *J Clin Psychiatry.* 2000;61:656-63. [PMID: 11030486]

Ballus C, Quiros G, De Flores T, de la Torre J, Palao D, Rojo L, et al. The efficacy and tolerability of venlafaxine and paroxetine in outpatients with depressive disorder or dysthymia. *Int Clin Psychopharmacol.* 2000;15:43-8. [PMID: 10836286]

McPartlin GM, Reynolds A, Anderson C, Casoy J. A comparison of oncedaily venlafaxine XR and paroxetine in depressed outpatients treated in general practice. *Primary Care Psychiatry.* 1998;4:127-32.

Coleman CC, Cunningham LA, Foster VJ, Batey SR, Donahue RM, Houser TL, et al. Sexual dysfunction associated with the treatment of depression: a placebo-controlled comparison of bupropion sustained release and sertraline treatment. *Ann Clin Psychiatry.* 1999;11:205-15. [PMID: 10596735]

Croft H, Settle E Jr, Houser T, Batey SR, Donahue RM, Ascher JA. A placebo-controlled comparison of the antidepressant efficacy and effects on sexual functioning of sustained-release bupropion and sertraline. *Clin Ther.* 1999;21: 643-58. [PMID: 10363731]

Kavoussi RJ, Segraves RT, Hughes AR, Ascher JA, Johnston JA. Double-blind comparison of bupropion sustained release and sertraline in depressed outpatients. *J Clin Psychiatry.* 1997;58:532-7. [PMID: 9448656]

Behnke K, Sogaard J, Martin S, Bauml J, Ravindran AV, Agren H, et al. Mirtazapine orally disintegrating tablet versus sertraline: a prospective onset of action study. *J Clin Psychopharmacol.* 2003;23:358-64. [PMID: 12920411]

Mehtonen OP, Sogaard J, Roponen P, Behnke K. Randomized, doubleblind comparison of venlafaxine and sertraline in outpatients with major depressive

disorder. Venlafaxine 631 Study Group. *J Clin Psychiatry*. 2000;61:95-100.  
[PMID: 10732656]